University of Thessaly
School of Engineering
Department of Mechanical Engineering

PhD Thesis

OPTIMIZATION OF FLEXIBLE PRODUCTION AND SUPPLY SYSTEMS

by

IOANNIS PAPACHRISTOS
Diploma in Mechanical Engineering, University of Thessaly, 2013
M.Sc., Department of Mechanical Engineering, University of Thessaly, 2014

A Thesis
Submitted in Partial Fulfilment of the
Requirements for the Degree of
Doctor of Philosophy
2019

Doctoral Thesis Examination Committee:

First Member (Supervisor)

Associate Professor Dimitrios Pandelis
Department of Mechanical Engineering, University of Thessaly

Second Member

Professor George Liberopoulos
Department of Mechanical Engineering, University of Thessaly

Third Member

Professor Epaminondas Kyriakidis
Department of Statistics, Athens University of Economics and Business

Fourth Member

Professor Athanasios Ziliaskopoulos
Department of Mechanical Engineering, University of Thessaly

Fifth Member

Professor Apostolos Burnetas
Department of Mathematics, University of Athens

Sixth Member

Associate Professor George Kozanidis
Department of Mechanical Engineering, University of Thessaly

Seventh Member

Associate Professor Stylianos Koukoumialos
General Department, University of Thessaly

Acknowledgements

The research reported in the thesis was carried out at the Laboratory of Production Organization and Industrial Management under the doctoral program of the Department of Mechanical Engineering of the University of Thessaly (UTH). The thesis work was funded by fellowships from the Department of Mechanical Engineering and by the Electronic Components and Systems for European Leadership (ECXEL) Joint Undertaking under Grant Agreement 737459.

First and foremost i would like to express my sincere gratitude to my thesis advisor, Associate Professor Dimitrios Pandelis who was abundantly helpful and offered invaluable assistance, support and guidance. I would like to thank him for the knowledge that he transmitted to me all these years, as well as for his influence in my way of thinking.

I would like to express my gratitude to Professor George Liberopoulos and Professor Epaminodas Kyriakidis for their continuous support. I gratefully thank Professor Athanasios Ziliaskopoulos, Professor Apostolos Burnetas, Associate Professor George Kozanidis and Associate Professor Stylianos Koukoumialos for accepting to be members of my thesis committee.

Last but not the least, i would like to thank my family: my parents, my sister and my fiancee for their support throughout writing this thesis.

Ioannis Papachristos

March 2019, Volos.

Abstract

Optimization of Flexible Production and Supply Systems

Ioannis Papachristos

Supervisor: Associate Professor Dimitrios Pandelis

In this thesis we deal with stochastic optimization problems that are related to the design and operation of flexible production and supply systems. In particular, we study server allocation problems in tandem queueing systems and the use of backup supply sources to hedge against supply risks.

For the first category of problems we consider two-stage queueing systems with one dedicated server in each station and a flexible server that can serve both stations. Assuming exponential service times and linear holding costs accrued by jobs present in the system, we seek optimal server allocation strategies within the classes of preemptive and non-preemptive policies for systems without external arrivals (clearing systems) and systems with Poisson arrivals under the discounted and the average cost criteria. For the model with a preemptive service discipline we assume that two servers can collaborate to work on the same job. When the combined rate of collaborating servers is less than the sum of their individual rates (partial collaboration), we identify conditions under which the optimal server allocation strategy is non-idling and has a threshold-type structure. Our results extend previous work on systems with additive service rates, either clearing or systems with arrivals and no dedicated server upstream. When the aforementioned conditions are not satisfied we show by examples that the optimal policy may have counterintuitive properties, which is not the case when a fully collaborative service discipline is assumed. We also obtain novel results for any type of collaboration when idling policies may be optimal and for systems with arrivals and dedicated servers in both stages. For the model with a non-preemptive service discipline we assume that the servers cannot collaborate and the dedicated servers are faster than the flexible server. We show that the dedicated server of the downstream station should never idle, and the same is true for the dedicated server of the upstream station when holding costs are larger there. On the other hand, the optimal allocation of the slow server is investigated through extensive numerical experiments that lead to conjectures on the structure of the optimal policy.

For the second category of problems we consider newsvendor models in which a retailer facing random demand with known distribution places an order to a primary supplier who

may not deliver the whole quantity ordered. We study two models of supply risk: suppliers who deliver a random portion of the order (random yield) and suppliers subject to random capacity, in which case the delivered quantity is limited by the realized capacity. To mitigate against such supply risks, the retailer contracts with a reliable backup supplier to buy the option to use his capacity after the delivery from the primary supplier. Depending on the responsiveness of the backup supplier, this option may be exercised before or after the demand becomes known as well. For the random yield case we also study models with two primary suppliers or two products sharing the same backup supplier. For all the aforementioned models we derive expressions for the optimal order and reservation quantities and obtain properties of these quantities. For the random capacity models we also determine the impact of the cost and revenue parameters on the optimal solution. Finally, we supplement our theoretical results with conjectures based on numerical experiments.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Efficient operations are required for a firm to be able to compete in a global environment. Malecki [52] cites three sectors of a firms operations that are crucial to the achievement of its goals: (i) inter-firm relations, (ii) resources and infrastructure, and (ii) workforce. In this dissertation we study models that are related to the three aforementioned sectors. Specifically, we consider server allocation problems in queueing networks, which are used to model production systems (sectors (ii) and (iii)), and newsvendor problems with multiple suppliers, which are used to model supply chain management issues (sector (i)). There are two unifying factors for these two classes of problems. First, our objective is to determine optimal rules for allocating resources to different tasks and for placing orders to the suppliers. Second, there are flexibility elements in our models; servers that can perform multiple tasks (e.g., cross-trained workforce) and backup suppliers for responding to supply shortages.

Due to the complexity of realistic production systems, the optimization of the queueing neworks that are used to model them is practically impossible. Therefore, in most cases the best we could expect is to come up with suboptimal policies that result in satisfactory performance with respect to some criterion (see, for example, Parvin et al. [63]). On the other hand, when we are concerned with optimality issues, we can either derive structural properties of optimal policies or study simple models for which we can determine optimal strategies. In this dissertation we consider two-stage queueing systems with one dedicated server for each stage and a flexible server that is trained to perform the tasks of both stages. We study variants of such systems resulting from whether preemptions are allowed or not and from various degrees of server collaboration. We address the questions of which stage should be given priority by the flexible server and to which server should a job be assigned in case collaboration is not permitted. By dealing with these optimization issues we get a deeper understanding of fundamental issues related to system performance, and insights that can be essential for performance improvement by forming the basis for the construction of good suboptimal policies.

Supply uncertainty is a problem often faced by firms and occurs when the delivered quantity is less than the quantity ordered, causing stock-outs and lost sales. The delivered quantity may be a fraction of the order (random yield) or may be constrained by the suppliers capacity (random capacity), which includes complete supply disruptions as an extreme special case. Examples of random yield causes include damages during transportation and uncertain production processes at the suppliers side. For instance, high-tech industries such as the semiconductor and liquid crystal display industries are known to experience high yield losses (Nahmias [54], Hu et al. [40]). On the other hand, random capacity is often associated with offshore suppliers who are characterized by relatively low reliability in terms of product delivery and quality (Sting and Huchzermeier [75]). In this dissertation we study random yield and random capacity models in a newsvendor context. As a supply risk mitigation strategy we consider the use of a reliable backup supplier whose capacity needs to be reserved in advance by paying a premium. Then, the firm has the option to order from the backup supplier after the supply uncertainty is resolved and either before or after the demand becomes known as well. For all the aforementioned models we obtain properties of the order and reservation quantities that maximize the firms expected profit.

The dissertation is organized as follows. Chapter 2 considers the server allocation problems in two-stage queueing systems. In Chapter 3 we analyze the newsvendor models. Long proofs are contained in appendices at the end of the manuscript. Note also that the material of Section 2.3.1 is under publication as an individual research paper (Papachristos and Pandelis [60]).

Chapter 2

Optimal Server Allocation in Two-Stage Queueing Systems

2.1 Introduction

We study two-station tandem queueing systems with one dedicated server in each station and one flexible server that is trained to work in both stations. Our objective is to determine properties of server allocation strategies that minimize expected linear holding costs for systems with exponential service times. The problem we consider is motivated by the use of cross-trained workers in manufacturing systems in order to cope with variability in demand, processing times, and operating conditions. Unlike traditional settings where each worker could perform a single task, cross-trained workers can be assigned to tasks where they are needed the most resulting in increased efficiency in the form of higher throughput, lower inventory, etc. Hopp and Van Oyen [35] have provided a literature survey on workforce flexibility as well as a framework for evaluating a flexible workforce in an organization. A more recent survey can be found in Andradottir et al. [11] along with design guidelines for eliminating bottlenecks.

We analyze systems with Poisson arrivals and systems without arrivals (clearing systems). Systems with arrivals model manufacturing facilities with continuous production where we are concerned with the long term performance of the system. On the other hand, some examples of clearing systems that occur in practice are the following: (i) production systems during end-of-shift operations where all unfinished work has to be completed, (ii) service systems (for example, banks) where no new customers are accepted after a certain time, but all customers already in the system have to be served, and (iii) production systems where at the beginning of each period (for example, every week) priority is given to unfinished work from previous periods, that is, new work orders have to wait until all previous orders are processed.

The models we study are also differentiated with respect to service discipline as we seek optimal server allocations within the classes of preemptive and nonpreemptive policies. In the first case a server may be reassigned to a different job at the time of an arrival or a service completion by another server. For this class of problems we extend previous work for various types of server collaboration. In the case of nonpreemptive policies a server must finish the processing of a job before being assigned to another job. Assuming that servers cannot collaborate to work on the same job and that the flexible server is slower than each dedicated server, we obtain some structural properties of the optimal policy and provide conjectures based on numerical experiments.

2.2 Related literature

Models of serial systems with the objective of maximizing throughput have received a lot of attention. For production lines with workers trained for all tasks (full cross-training) Van Oyen et al. [81] computed the improvement in throughput that can be achieved by worker flexibility as opposed to the optimal static allocation. Hopp et al. [36] demonstrated the effectiveness of D-skill chains where each worker is trained for his base station and $D-1$ more tasks down the line (U-shaped lines were assumed) with emphasis on 2-skill chains. Parvin et al. [63] presented a zone chaining pattern with limited cross-training that can achieve high throughput. Finally, for models limited to 2 or 3 stages, throughput maximizing policies were determined by Andradottir et al. [8],[9],[10],[12], Andradottir and Ayhan [7], Gel et al. [29], Arumugam et al. [14], Hasenbein and Kim [33], Kirkizlar et al. [44], and Wang et al. [83].

Because of the complexity of the mathematical models involved, research on the optimal use of flexible servers with holding costs has focused on two-stage Markovian systems. Rosberg et al. [64] considered a system with Poisson arrivals, a server with a constant service rate in the downstream station, and a server with controllable service rate in the upstream station. They showed that the optimal service rate is non-decreasing in the length of the first queue and non-increasing in the length of the second queue. For a clearing system with two flexible servers, Ahn et al. [3] provided necessary and sufficient conditions under which an exhaustive policy for the upstream or the downstream station is optimal. Similar results were obtained by Ahn et al. [4] for the model with arrivals. The results of Ahn et al. [3] have been extended in two directions. First, Schiefermayr and Weichbold [67] obtained the optimal policy for all values of holding costs and service times, and second, Weichbold and Schiefermayr [86] derived conditions for the optimality of exhaustive policies when jobs require the second stage of service with a certain probability. Kirkizlar et al. [45] considered a problem where, in addition to holding costs, a profit is earned whenever a job is completed. For a tandem system with two flexible servers they showed that the profit maximizing strategy is characterized by a threshold and determined the value of this threshold.

A common characteristic of the models studied in the aforementioned papers (with the exception of Rosberg et al. [64]) is that they did not include dedicated servers. Farrar [22],[23] considered two versions of a clearing system with dedicated servers in each station and one flexible server. In the constrained version the flexible server can only work in the upstream station, whereas in the unconstrained version the server can work in both stations. He showed that for both versions the optimal policy is characterized by a switching curve; the flexible server is idled or assigned to the downstream station if the number of jobs there exceeds a threshold that depends on the number of jobs in the first queue. He also showed that the slope of the switching curve is at least -1, indicating that if the flexible server is idled or assigned to the downstream station, its allocation does not change if a job joins the queue from upstream (*transition monotone* policy). Pandelis [56] extended the results of Farrar [22],[23] to the case when jobs may leave the system after completing service in the first station and specified subsets of the state space where the optimal policy can be explicitly determined. The same structure of the optimal policy was obtained by Wu et al. [87] where it was assumed that the servers have varying speeds and the processing requirements are the same in both stations. Wu et al. [88] showed the optimality of a switching-curve policy for the previous model with arrivals and no dedicated server in the upstream station and Pandelis [57] extended this result to the case when jobs may not require service at the downstream station and processing requirements are not the same in each station. Finally, Pandelis [58] studied a model with server operating costs in addition to holding costs and identified conditions under which the switching-curve structure of the optimal policy is preserved. With the exception of Pandelis [56] (constrained version), a common assumption in all of the aforementioned papers was that different servers could collaborate to work on the same job, in which case the total service rate was equal to the sum of the individual servers rates (*fully collaborative* servers). Moreover, a non-idling discipline for at least the dedicated servers was assumed. Both of these conditions were relaxed by Pandelis [59]. For clearing systems he showed that non-idling policies are optimal when the holding cost rate in the upstream station is not less than the corresponding rate in the downstream station, and for this case he provided conditions on service rates that ensure that the optimal server allocation is characterized by a single switching curve under a non-collaborative service discipline.

With regard to nonpreemptive policies, to the best of our knowledge there is no previous work on two-stage systems with both dedicated and flexible servers. There has been a lot of attention to one-stage systems where a stream of jobs is served by a fast and a slow server, which is known as the slow server problem. The optimal policy for this problem dictates that the fast server should not idle and the slow server should be used when the number of jobs exceeds a certain threshold. This result has been proved by Lin and Kumar [51], Walrand [82], Stockbridge [76], Xu [90], and Koole [48]. The threshold-type property of the optimal policy has also been shown for models with operating costs (Akgun et al. [6]) and servers subject to failures (Ozkan and Kharoufeh [55]). For problems with more than two servers and no arrivals, Agrawala et al. [1] proved that the optimal policy is determined by multiple thresholds. This is still an open problem for systems with arrivals. Rosberg and Makowski [65] showed that the aforementioned multiple threshold policy is optimal for sufficiently small arrival rates and Weber [85] discussed the conjecture that this policy is optimal for arbitrary arrival rates.

2.3 Preemptive service discipline

In this section we focus on systems where two servers can collaborate to work on the same job but their combined service rate is less than the sum of their individual service rates (*partially collaborative* servers). Situations like this arise when for some reason (e.g., servers sharing resources when collaborating) it is not possible for each server to achieve full performance. The assumption of non-additive service rates has also been used in the work of Ahn and Lewis [5] who studied the problem of optimal routing and flexible server allocation to two parallel queues. In addition to partially collaborating servers (subadditive rates) they considered the case when collaboration increases the servers' efficiency, that is, their combined service rate is larger than the sum of their individual service rates (superadditive rates). Models with non-additive rates for tandem systems with throughput maximization as the objective were studied by Andradottir et al. [12] (subadditive rates), Andradottir et al. [10] and Wang et al. [83] (superadditive rates). In the context of tandem systems with dedicated servers in each station it will become evident from the analysis that the problem with superadditive service rates is equivalent to a problem with fully collaborative servers, so we do not consider this case.

In the following section we study clearing systems. For the case of non-idling optimal policies we extend results from past literature by providing conditions on service rates under which the structure of the optimal policy for fully collaborative servers is preserved under partial collaboration. When these conditions are not satisfied we show by examples that the optimal server allocation may not possess the same structure and in fact be quite counterintuitive. When idling policies are optimal we obtain properties of the optimal policy that are novel for any type of collaboration. Specifically, we provide an asymptotic characterization of the optimal policy for a large number of jobs in the downstream station, and in case of no dedicated server in one of the stations we show that the optimal allocation is determined by a switching curve. In Section 2.3.2 we study systems with arrivals under the discounted and average cost criteria. For systems with one dedicated server we extend to the partial collaboration case some of the results obtained by Wu et al. [88] and Pandelis [57] for fully collaborative servers. Furthermore, for systems with dedicated servers in both stations and any type of collaboration, we explicitly determine the discounted cost optimal policy for a subset of the state space.

2.3.1 Clearing systems

We study two-stage tandem queueing systems with a number of jobs initially present and no further arrivals. After their service is completed in the upstream station (Station 1), jobs move to the downstream station (Station 2) where they receive additional service, and then they leave the system. Each job in station i, $i = 1, 2$, incurs linear holding costs at rate h_i. There are dedicated servers, one for each station, that are trained to work only in their corresponding station, and one flexible server that can work in both stations. We assume that this server can move from station to station instantaneously without any cost. We assume exponential service times with rates ν_1, ν_2 for jobs served by the dedicated server and μ_1, μ_2 for jobs served by the flexible server in Station 1,2, respectively. We assume that two servers can work simultaneously on different jobs in the same station, as well as collaborate to work on the same job. When the collaboration takes place in Station i, $i = 1, 2$, the service rate is equal to $\nu_i + \xi_i$, where $\nu_i + \xi_i > \mu_i$ and $0 < \xi_i \leq \mu_i$, with equality corresponding to full collaboration. Our objective is to find a server allocation strategy that minimizes the total expected holding cost until the system is cleared of all jobs.

We formulate the problem as a Markov decision process with state space $\{(x_1, x_2) : x_1, x_2 \geq 0\}$, where x_i, $i = 1, 2$, is the number of jobs in Station i, including those in service. Starting from state (x_1, x_2), we denote by $V(x_1, x_2)$ the minimum total expected holding cost until the system empties, with $V(0,0)=0$. Instead of the continuous time problem, we study an equivalent discrete time problem obtained by uniformization (see, e.g., [74]), where without loss of generality we assume $\nu_1+\nu_2+\mu_1+\mu_2+\xi_1+\xi_2=1$. Then, with $A(x_1,x_2)$ denoting the set of feasible service rates in state (x_1,x_2), we get the following optimality equation.

$$V(x_1,x_2)=h_1x_1+h_2x_2+\min_{(\rho_1,\rho_2)\in A(x_1,x_2)} W_{\rho_1,\rho_2}(x_1,x_2), \tag{2.3.1}$$

where

$$W_{\rho_1,\rho_2}(x_1,x_2)=\rho_1V(x_1-1,x_2+1)+\rho_2V(x_1,x_2-1)+(1-\rho_1-\rho_2)V(x_1,x_2). \tag{2.3.2}$$

Note that if $x_1=0$ (resp. $x_2=0$), we get $V(-1,x_2+1)$ (resp. $V(x_1,-1)$) in (2.3.2), which are terms that have not been formally defined. However, this is not a problem because the only feasible rate is $\rho_1=0$ (resp. $\rho_2=0$).

Before proceeding to the characterization of the optimal policy, we give preliminary results that will be used in the proof of the main results of this section. Lemma 2.1 states that the minimum expected cost increases with the number of jobs in the system.

Lemma 2.1. *$V(x_1,x_2)$ is increasing in x_1 and x_2.*

Lemma 2.2 gives an auxiliary result that will be used in comparisons that determine the optimal server allocation.

Lemma 2.2. *Suppose that $A-B=G+\alpha(A-B)+\beta(A^- - B^-)+\gamma(A^+ - B^+)$, where $A,B,G,\alpha,\beta,\gamma$ are real numbers with $\alpha+\beta<1$ and $\alpha+\gamma<1$. Then, $A-B$ and G have the same sign.*

Proof. Assume $A<B$. Then

$$\begin{aligned} A<B\le 0 &\implies G=(1-\alpha-\beta)(A-B)<0,\\ A<0\le B &\implies G=(1-\alpha-\beta)A-(1-\alpha-\gamma)B<0,\\ 0\le A<B &\implies G=(1-\alpha-\gamma)(A-B)<0. \end{aligned}$$

By the same reasoning $A>B$ implies that $-G<0$, and the proof is complete. □

2.3.1.1 The optimal policy

First, when one of two queues is empty of jobs, it is clear that the optimal policy allocates the maximum possible service rate to the nonempty queue. When there is one job, the dedicated and the flexible server work together on that job, otherwise they work on separate jobs. Therefore,

$$V(1,0) = \frac{h_1}{\nu_1+\xi_1}+V(0,1), \tag{2.3.3}$$

$$V(x_1,0) = \frac{h_1x_1}{\nu_1+\mu_1}+V(x_1-1,1), \quad x_1>1 \tag{2.3.4}$$

$$V(0,1) = \frac{h_2}{\nu_2+\xi_2}+V(0,0), \tag{2.3.5}$$

$$V(0,x_2) = \frac{h_2x_2}{\nu_2+\mu_2}+V(0,x_2-1), \quad x_2>1. \tag{2.3.6}$$

We consider now the optimal allocation in the downstream station. It is reasonable to expect that the optimal policy would allocate as much service rate as possible to Station 2 to push jobs out of the system, thus saving holding costs. To prove this formally, we define function $g(x_1, x_2)$ as

$$g(x_1, x_2) = V(x_1, (x_2 - 1)^+) - V(x_1, x_2), \quad x_1, x_2 \geq 0.$$

Assuming an initial allocation ρ_1, ρ_2 for some state with $x_2 \geq 1$, the incentive to allocate additional rate ρ to Station 2 is equal to

$$W_{\rho_1,\rho_2}(x_1, x_2) - W_{\rho_1,\rho_2+\rho}(x_1, x_2) = -\rho g(x_1, x_2)$$

by (2.3.2). This incentive is positive because of Lemma 2.1, leading to

Proposition 2.1. *For given ρ_1, $W_{\rho_1,\rho_2}(x_1, x_2)$ is minimized by maximizing ρ_2.*

A consequence of Proposition 2.1 is that the optimal policy does not idle the dedicated server in Station 2. Turning to the optimal allocation in the upstream station, the incentive to allocate additional rate ρ to Station 1 is equal to $\rho f(x_1, x_2)$, where

$$f(x_1, x_2) = V(x_1, x_2) - V(x_1 - 1, x_2 + 1), \quad x_1 \geq 1, \ x_2 \geq 0.$$

Therefore, we obtain the following proposition.

Proposition 2.2. *For given ρ_2, $W_{\rho_1,\rho_2}(x_1, x_2)$ is minimized by maximizing ρ_1 if $f(x_1, x_2) \geq 0$ and by $\rho_1 = 0$ if $f(x_1, x_2) < 0$.*

Proposition 2.2 indicates that, depending on the sign $f(x_1, x_2)$, the optimal policy should either allocate as many resources as possible to Station 1 or no resources at all. Taking into account Proposition 2.1 as well, we conclude that the optimal policy does not idle the dedicated servers when $f(x_1, x_2) \geq 0$, whereas in the opposite case it idles the dedicated server of Station 1 and assigns the flexible server to Station 2 to work along with its dedicated server.

Remark 2.1. *Propositions 2.1 and 2.2 hold for any form of collaboration. Then, assuming superadditive service rates, that is, $\xi_i > \mu_i$, $i = 1, 2$, the optimal policy would always have the flexible server collaborating with one of the dedicated servers, say server i, to work on the same job, resulting in a total service rate of $\nu_i + \xi_i$. This is equivalent to an additive service rate model with rates ξ_1, ξ_2 for jobs served by the flexible server in Station 1 and 2, respectively.*

2.3.1.2 Optimality of non-idling policies

In this section we investigate the structure of optimal policies for $h_1 \geq h_2$, which is a necessary and sufficient condition for the optimality of non-idling policies. Intuitively, when it is not cheaper to have jobs in Station 1 compared to Station 2, it is reasonable not to idle resources to keep jobs upstream. The necessity of the condition is proved in the next section (Theorem 2.6). The sufficiency is a consequence of the following lemma.

Lemma 2.3. *Let $h_1 \geq h_2$. Then, $f(x_1, x_2) > 0$ for all $x_1 \geq 1$, $x_2 \geq 0$.*

Because idling a server cannot be optimal, the decision to be made is where to assign the flexible server. Propositions 2.1 and 2.2 also imply that when there are at least two jobs in the station to which the flexible server is assigned, the two servers should work on separate jobs rather than collaborate on the same job. Therefore, taking also into account that $\nu_1 + \nu_2 + \mu_1 + \mu_2 + \xi_1 + \xi_2 = 1$, the optimality equations take the following form:

$$\begin{aligned} V(1,1) &= h_1 + h_2 + \nu_1 V(0,2) + \nu_2 V(1,0) + (\mu_1 + \mu_2) V(1,1) \\ &\quad + \min\{\xi_1 V(0,2) + \xi_2 V(1,1), \xi_2 V(1,0) + \xi_1 V(1,1)\}, \end{aligned} \tag{2.3.7}$$

and for $x_1, x_2 > 1$,

$$\begin{aligned} V(1,x_2) &= h_1 + h_2 x_2 + \nu_1 V(0,x_2+1) + \nu_2 V(1,x_2-1) + (\xi_2 + \mu_1) V(1,x_2) \\ &\quad + \min\{\xi_1 V(0,x_2+1) + \mu_2 V(1,x_2), \mu_2 V(1,x_2-1) + \xi_1 V(1,x_2)\}, \end{aligned} \tag{2.3.8}$$

$$\begin{aligned} V(x_1,1) &= h_1 x_1 + h_2 + \nu_1 V(x_1-1,2) + \nu_2 V(x_1,0) + (\xi_1 + \mu_2) V(x_1,1) \\ &\quad + \min\{\mu_1 V(x_1-1,2) + \xi_2 V(x_1,1), \xi_2 V(x_1,0) + \mu_1 V(x_1,1)\}, \end{aligned} \tag{2.3.9}$$

$$\begin{aligned} V(x_1,x_2) &= h_1 x_1 + h_2 x_2 + \nu_1 V(x_1-1,x_2+1) + \nu_2 V(x_1,x_2-1) \\ &\quad + (\xi_1 + \xi_2) V(x_1,x_2) + \min\{\mu_1 V(x_1-1,x_2+1) + \mu_2 V(x_1,x_2), \\ &\quad \mu_2 V(x_1,x_2-1) + \mu_1 V(x_1,x_2)\}, \end{aligned} \tag{2.3.10}$$

where the first and second terms in braces correspond to the assignment of the flexible server to the first and second station, respectively. Next, we define a set of functions that characterize the optimal decision in each state.

$$d(x_1,x_2) = \mu_1 f(x_1,x_2) + \mu_2 g(x_1,x_2), \quad x_1 \geq 1,\ x_2 \geq 0, \tag{2.3.11}$$

$$\tilde{d}(1,x_2) = \xi_1 f(1,x_2) + \mu_2 g(1,x_2), \quad x_2 \geq 0, \tag{2.3.12}$$

$$\hat{d}(x_1,1) = \mu_1 f(x_1,1) + \xi_2 g(x_1,1), \quad x_1 \geq 1, \tag{2.3.13}$$

$$\bar{d}(1,1) = \xi_1 f(1,1) + \xi_2 g(1,1). \tag{2.3.14}$$

Function $d(x_1, x_2)$ is derived by subtracting the first from the second term in curly brackets in (2.3.10). Therefore, its sign determines the optimal allocation for the flexible server when there are at least two jobs in each station: assign the server upstream if $d(x_1, x_2) \geq 0$, and downstream otherwise. Similarly, $\hat{d}(x_1, 1)$ is the decision function when there is one job in the downstream station and at least two jobs upstream, $\tilde{d}(1, x_2)$ is the decision function when there is one job in the upstream station and at least two jobs downstream, and $\bar{d}(1, 1)$ is the decision function when there is one job in each station.

In the main result of this section given in Theorem 2.1 we give conditions under which properties of the optimal policy that have been shown to hold for fully collaborative servers also hold for partially collaborative servers. The proof of the theorem requires the following three lemmas.

Lemma 2.4. *Let $\nu_2 \geq \mu_2$. Then*
i) For $x_2 \geq 0$,

$$\tilde{d}(1, x_2 + 1) < \tilde{d}(1, x_2).$$

ii) $\lim_{x_2 \to \infty} \tilde{d}(1, x_2) = -\infty$.

Lemma 2.5. *Let $\nu_2 \geq \mu_2$ and $\mu_1 \geq \mu_2$. Then*
i) For $x_1 \geq 1$, $x_2 \geq 0$,

$$d(x_1, x_2 + 1) < d(x_1, x_2).$$

ii) For $x_1 \geq 1$

$$\lim_{x_2 \to \infty} d(x_1, x_2) = -\infty.$$

Lemma 2.6. *Let $\nu_2 \geq \mu_2$ and $\mu_1 \geq \mu_2$. Then*
i) For $x_2 \geq 1$,

$$d(2, x_2) \geq 0 \Longrightarrow d(2, x_2) \geq d(1, x_2 + 1).$$

ii) For $x_2 \geq 2$,

$$\tilde{d}(1, x_2) \geq 0 \Longrightarrow \tilde{d}(1, x_2) \leq d(2, x_2 - 1).$$

iii) For $x_1 \geq 2$, $x_2 \geq 2$,

$$d(x_1, x_2) \geq 0 \Longrightarrow d(x_1, x_2) \leq d(x_1 + 1, x_2 - 1).$$

Lemmas 2.4 and 2.5 are used to prove that the optimal policy is determined by a single switching curve, formally defined in Theorem 2.1 and Lemma 2.6 to obtain a lower bound on its slope.

Theorem 2.1. *Assume $h_1 \geq h_2$, $\nu_2 \geq \mu_2$, $\mu_1 \geq \mu_2$, and partially collaborative servers. Then, for each $x_1 \geq 1$, there exists an integer $t(x_1) \geq 1$ such that the optimal policy assigns the flexible server to Station 2 (resp. 1) when $x_2 \geq t(x_1)$ (resp. $x_2 < t(x_1)$). Moreover, the slope of $t(x_1)$ is at least -1.*

Proof. To prove the existence part, we first consider $x_1 = 1$. If $\bar{d}(1, 1) < 0$, then for $x_2 \geq 2$ we have $\tilde{d}(1, x_2) < 0$ because of $\tilde{d}(1, 1) < \bar{d}(1, 1)$ and part (i) of Lemma 2.4. Therefore, the optimal policy assigns the flexible server to the downstream station for any number of jobs in that station, that is, $t(1) = 1$. Otherwise, let $m = \min\{x_2 \geq 2 : \ \tilde{d}(1, x_2) < 0\}$, noting that the existence of this minimum is guaranteed by part (ii) of Lemma 2.4. Then, $t(1) = m$ because part (i) of Lemma 2.4 implies $\tilde{d}(1, x_2) < 0$ for $x_2 \geq m$. For $x_1 > 1$ the statement of the theorem is proved similarly by using the fact that $d(x_1, 1) < \hat{d}(x_1, 1)$ and Lemma 2.5.

The fact that the slope of $t(x_1)$ is at least -1 is a consequence of parts (ii) and (iii) of Lemma 2.6, from which it follows that if the decision function is negative at some state (x_1, x_2), it is also negative at $(x_1 - 1, x_2 + 1)$. □

As seen from its statement, we were able to prove Theorem 2.1 under conditions $\nu_2 \geq \mu_2$ and $\mu_1 \geq \mu_2$. The first one implies that the specialist (dedicated server) in Station 2 is not slower than the generalist (flexible) server, which is a reasonable assumption. However, this is not the case with the second condition which seems arbitrary. An interesting question is whether the two conditions are crucial for the validity of the results, or they were just needed for the arguments of the proofs to work. To answer this question we obtained numerical results that illustrate the structure of the optimal policy when either one or both of the conditions are violated. For each of the three cases we created 100,000 problem instances with randomly generated values for service and holding cost rates and computed the optimal server allocation for each one. When only one of the conditions was violated, all of our results were in agreement with Theorem 2.1. Moreover, we observed that the switching curve was nondecreasing in all instances. When both conditions were violated, the optimal policy was still determined by a unique switching curve, but we found instances with switching curves having a portion with slope less than -1. One such instance is given in the following example.

Example 2.1. *Let* $\nu_1 = 0.8$, $\mu_1 = 0.6$, $\xi_1 = 0.03$, $\nu_2 = 0.6$, $\mu_2 = 8$, $\xi_2 = 7.43$, $h_1 = 16$, *and* $h_2 = 1.5$. *When there are three jobs in each station, the optimal policy assigns the flexible server to Station 2. However, if a job completes its service in Station 1 and joins Station 2, then, contrary to intuition, the flexible server is transferred to Station 1 (Figure 2.1).*

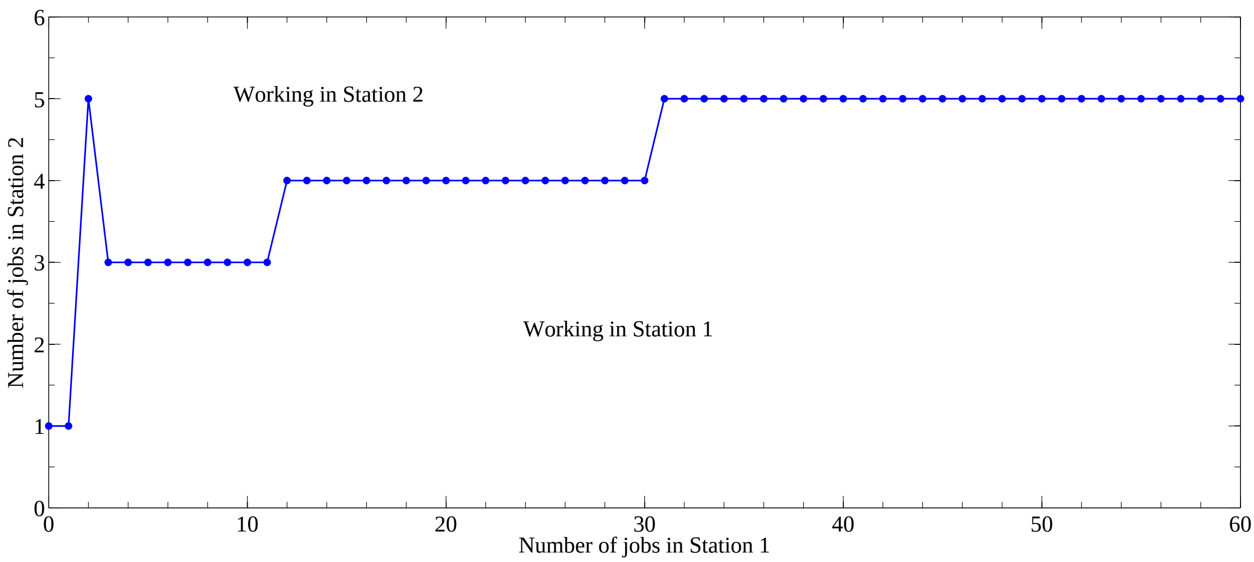

Figure 2.1: Switching points for the flexible server

When collaboration is not allowed in Station 1, we can prove that the optimal policy is characterized by a single switching curve without condition $\mu_1 \geq \mu_2$ (Theorem 2.2). First, we prove the following lemma.

Lemma 2.7. *Let* $\xi_1 = 0$, $\nu_1 \geq \mu_1$, *and* $\nu_2 \geq \mu_2$. *Then*
i) For $x_2 \geq 0$,

$$f(1, x_2 + 1) < f(1, x_2).$$

ii) For $x_1 \geq 1$, $x_2 \geq 0$,

$$d(x_1, x_2 + 1) < d(x_1, x_2).$$

iii) For $x_1 \geq 1$

$$\lim_{x_2 \to \infty} d(x_1, x_2) = -\infty.$$

Theorem 2.2. *Assume $h_1 \geq h_2$, $\nu_1 \geq \mu_1$, $\nu_2 \geq \mu_2$, non-collaborative servers in Station 1 and partially collaborative servers in Station 2. Then*
i) For $x_1 = 1$ the optimal policy assigns the flexible server to Station 2.
ii) For each $x_1 \geq 2$, there exists an integer $t(x_1) \geq 1$ such that the optimal policy assigns the flexible server to Station 2 (resp. 1) when $x_2 \geq t(x_1)$ (resp. $x_2 < t(x_1)$).

Proof. The first part is a direct consequence of Proposition 2.2 and Lemma 2.3. The proof of the second part is similar to the proof of the analogous part of Theorem 2.1 and is based on the monotonicity and asymptotic properties of decision function $d(x_1, x_2)$ established in parts (ii) and (iii) of Lemma 2.7. □

A special case of interest is when there is only one dedicated server. For this case the optimal policy is either characterized by a switching curve (as in Theorem 2.1) or it can be explicitly determined by a $c\mu$-type rule according to which the flexible server is assigned to the station with no dedicated server if the holding cost savings from a service completion in that station are not less than the corresponding savings from a service completion in the other station.

The following theorem gives properties of the optimal policy when there is no dedicated server assigned to Station 1. Note that in this case non-idling policies are optimal for any values of holding cost rates so condition $h_1 \geq h_2$ is not needed.

Theorem 2.3. *Assume $\nu_1 = 0$, $\nu_2 \geq \mu_2$, and partially collaborative servers in Station 2. Then*
i) When $\mu_1(h_1 - h_2) < \mu_2 h_2$, for each $x_1 \geq 1$, there exists an integer $t(x_1) \geq 1$ such that the optimal policy assigns the flexible server to Station 2 (resp. 1) when $x_2 \geq t(x_1)$ (resp. $x_2 < t(x_1)$). Moreover, the slope of $t(x_1)$ is at least -1.
ii) When $\mu_1(h_1 - h_2) \geq \mu_2 h_2$, the optimal policy assigns the flexible server to Station 1 for all $x_1 \geq 1$.

As with Theorem 2.1, we conducted an extensive numerical investigation to see whether condition $\nu_2 \geq \mu_2$ is needed for the validity of Theorem 2.3 by examining 100,000 test cases with $\nu_2 < \mu_2$ for each of the two parts of the theorem. We found the structure of the optimal policy for all of them to be in agreement with the theorem. In addition, $t(x_1)$ was nondecreasing for all cases.

When there is no dedicated server assigned to Station 2, the optimal policy is characterized in the following theorem.

Theorem 2.4. *Assume $h_1 \geq h_2$, $\nu_2 = 0$, $\mu_1 \geq \mu_2$, and partially collaborative servers in Station 1. Then*
i) For each $x_1 \geq 1$, there exists an integer $t(x_1) \geq 1$ such that the optimal policy assigns the flexible server to Station 2 (resp. 1) when $x_2 \geq t(x_1)$ (resp. $x_2 < t(x_1)$). Moreover, the slope of $t(x_1)$ is at least -1.
ii) When $\mu_1(h_1 - h_2) \leq \mu_2 h_2$, the optimal policy assigns the flexible server to Station 2, that is, $t(x_1) = 1$.

Similarly to Theorem 2.1, we used numerical experiments to examine the effect of condition $\mu_1 \geq \mu_2$ on the validity of Theorem 2.4. When the condition is violated, we found that part (i) holds for $\mu_1(h_1 - h_2) > \mu_2 h_2$. On the other hand, the $c\mu$-type rule implied by part (ii) may not be optimal when $\mu_1 < \mu_2$. This is illustrated in the following example.

Example 2.2. *Let $\nu_1 = 1.3$, $\mu_1 = 0.9$, $\xi_1 = 0.1$, $\mu_2 = 7.7$, $h_1 = 11.4$, and $h_2 = 1.2$, so that $\mu_1(h_1 - h_2) < \mu_2 h_2$. However, there are states (e.g., $x_1 = 3$, $x_2 = 1$) for which the flexible server is assigned to Station 1 (Figure 2.2).*

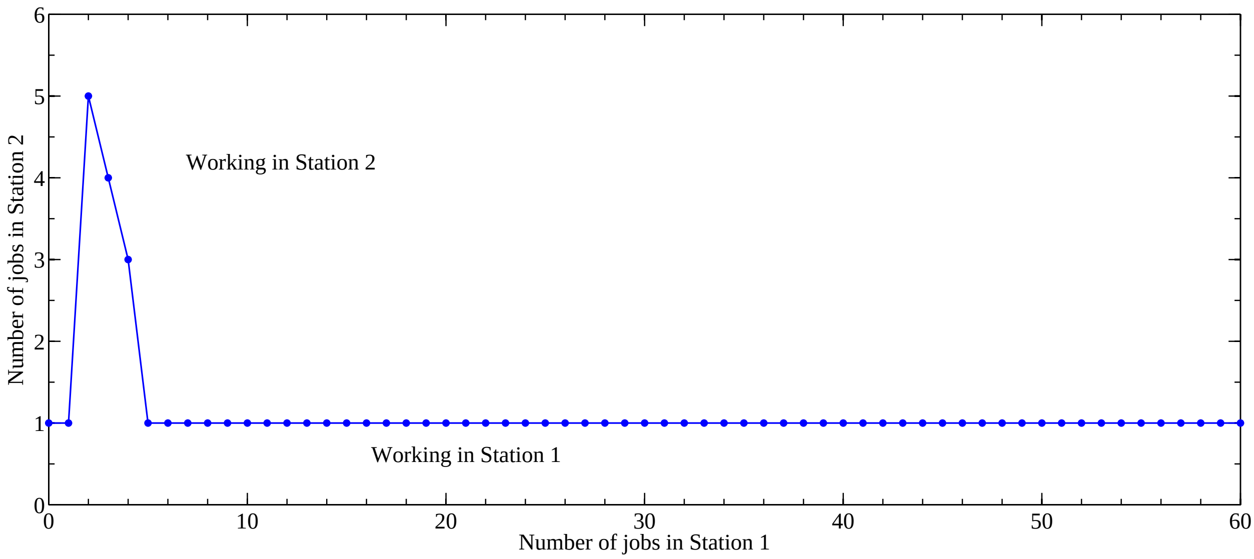

Figure 2.2: Switching points for the flexible server

When collaboration is not allowed in Station 1, we can prove that the optimal policy is characterized by a single switching curve without condition $\mu_1 \geq \mu_2$.

Theorem 2.5. *Assume $h_1 \geq h_2$, $\nu_1 \geq \mu_1$, $\nu_2 = 0$, and non-collaborative servers in Station 1. Then*
i) For $x_1 = 1$ the optimal policy assigns the flexible server to Station 2.
ii) For each $x_1 \geq 2$, there exists an integer $t(x_1) \geq 1$ such that the optimal policy assigns the flexible server to Station 2 (resp. 1) when $x_2 \geq t(x_1)$ (resp. $x_2 < t(x_1)$).

2.3.1.3 Optimality of idling policies

In this section we investigate the structure of optimal policies for $h_1 < h_2$. When jobs in Station 2 incur larger holding costs than jobs in Station 1, it may be optimal to keep jobs from joining Station 2 by not assigning any resources to Station 1, that is, there exist states (x_1, x_2) for which $f(x_1, x_2) < 0$. In the following theorem we show that this is indeed the case.

Theorem 2.6. *Assuming $h_1 < h_2$, for each $x_1 \geq 1$, there exists an integer $t(x_1) \geq 1$ such that $f(x_1, x_2) < 0$ for $x_2 \geq t(x_1)$.*

Proof. It suffices to show that $f(x_1, x_2)$ is decreasing in x_2 for x_2 sufficiently large and $\lim_{x_2 \to \infty} f(x_1, x_2) = -\infty$. The proof is by induction on x_1. For $x_2 \geq 1$ we have

$$\begin{aligned} f(1, x_2+1) &= V(1, x_2+1) - V(0, x_2+2) \\ &\leq h_1 + h_2(x_2+1) + W_{0,\nu_2+\mu_2}(1, x_2+1) - V(0, x_2+2) \\ &= h_1 - h_2 + (\nu_2+\mu_2)V(1, x_2) + (\nu_1+\mu_1+\xi_1+\xi_2)V(1, x_2+1) \\ &\quad -(\nu_2+\mu_2)V(0, x_2+1) - (\nu_1+\mu_1+\xi_1+\xi_2)V(0, x_2+2) \\ &= h_1 - h_2 + (\nu_2+\mu_2)f(1, x_2) + (\nu_1+\mu_1+\xi_1+\xi_2)f(1, x_2+1) \\ \implies & (\nu_2+\mu_2)\left[f(1, x_2+1) - f(1, x_2)\right] \leq h_1 - h_2 < 0, \end{aligned}$$

which proves the result for $x_1 = 1$ and establishes the induction base. Assume that the result holds for some $x_1 > 1$, which implies that there exists $t(x_1)$ such that the optimal allocation for $x_2 \geq \max(t(x_1), 2)$ is $(0, \nu_2 + \mu_2)$. Then, for $x_2 \geq \max\{t(x_1) - 2, 1\}$ we can replicate the arguments used for $x_1 = 1$ to show that

$$(\nu_2+\mu_2)\left[f(x_1+1, x_2+1) - f(x_1+1, x_2)\right] \leq h_1 - h_2 < 0,$$

which completes the induction and the proof. □

In the following theorem we show that when there is one job in Station 1 and the flexible server is not faster than the dedicated server in Station 2, the optimal policy is determined by two switching points.

Theorem 2.7. *Assume $h_1 < h_2$, $\nu_2 \geq \mu_2$, and $x_1 = 1$. Then, there exist integers $t_2 \geq t_1 \geq 1$ such that the optimal policy idles the dedicated server in Station 1 when $x_2 \geq t_2$ and assigns the flexible server to Station 2 (resp. 1) when $x_2 \geq t_1$ (resp. $x_2 < t_1$).*

Proof. The existence of t_2 follows from $f(1, x_2)$ being decreasing (see proof of Theorem 2.6). For $x_2 < t_2$ we have $f(x_1, x_2) \geq 0$ and optimality equations (2.3.7) and (2.3.8) hold. Therefore, the optimal allocation of the flexible server depends on the sign of $\tilde{d}(1, x_2)$, which is decreasing by Lemma 2.4(i), proving the existence of the lower switching point t_1. □

We believe that Theorem 2.7 is also valid for more than one job in Station 1, that is, for a fixed number of jobs in Station 1 the optimal policy is determined by two switching points $t_2(x_1) \geq t_1(x_1) \geq 1$. Our conjecture was verified by extensive numerical experiments but we were able to prove it only when there is no dedicated server assigned to Station 2. The optimal policy for this case is given in the following theorem.

Theorem 2.8. *Let $h_1 < h_2$ and $\nu_2 = 0$. Then*
i) The optimal policy assigns the flexible server to Station 2, that is, $t_1(x_1) = 1$.
ii) For each $x_1 \geq 1$, there exists an integer $t_2(x_1) \geq 1$ such that the optimal policy idles the dedicated server of Station 1 when $x_2 \geq t_2(x_1)$. Moreover, $t_2(x_1)$ is nondecreasing.

We end the section by pointing out that the results of Theorem 2.6, Theorem 2.7 (without condition $\nu_2 \geq \mu_2$) and Theorem 2.8 are novel for the case of fully collaborative servers as well.

2.3.2 Systems with arrivals

In this section we study the previous model with Poisson arrivals of rate λ and preemptions allowed at arrival times. We consider the discrete time equivalent problem by applying uniformization with rate $\lambda+\nu_1+\nu_2+\mu_1+\mu_2+\xi_1+\xi_2=1$. With β being a discount factor, where $0<\beta\leq 1$, we denote by $V^{\theta}_{n,\beta}(x_1,x_2)$ and $V^{\theta}_{\beta}(x_1,x_2)=\lim_{n\to\infty}V^{\theta}_{n,\beta}(x_1,x_2)$ the expected n-step discounted cost and the expected infinite horizon discounted cost, respectively, under policy θ starting at state (x_1,x_2). We also define the expected average cost under θ by,

$$J^{\theta}(x_1,x_2)=\limsup_{n\to\infty}\frac{V^{\theta}_{n,1}(x_1,x_2)}{n}.$$

Our objective is to derive structural properties of server allocation strategies that minimize the expected infinite horizon discounted cost and the expected average cost.

2.3.2.1 Discounted cost

Let $V_{n,\beta}(x_1,x_2)$ be the minimum expected n-step discounted cost (with $V_{0,\beta}(x_1,x_2)=0$), for which it is clear that it satisfies the following optimality equation.

$$V_{n,\beta}(x_1,x_2)=h_1x_1+h_2x_2+\beta\min_{(\rho_1,\rho_2)\in A(x_1,x_2)}T_{\rho_1,\rho_2}V_{n-1,\beta}(x_1,x_2), \tag{2.3.15}$$

where operator T_{ρ_1,ρ_2} is defined by

$$\begin{aligned} T_{\rho_1,\rho_2}U(x_1,x_2) &= \lambda U(x_1+1,x_2)+\rho_1U(x_1-1,x_2+1) \\ &\quad +\rho_2U(x_1,x_2-1)+(1-\lambda-\rho_1-\rho_2)U(x_1,x_2). \end{aligned} \tag{2.3.16}$$

Next, we denote by $V_{\beta}(x_1,x_2)=\inf_{\theta}V^{\theta}_{\beta}(x_1,x_2)$ the minimum expected infinite horizon discounted cost; $V_{\beta}(x_1,x_2)$ satisfies optimality equation

$$V_{\beta}(x_1,x_2)=h_1x_1+h_2x_2+\beta\min_{(\rho_1,\rho_2)\in A(x_1,x_2)}T_{\rho_1,\rho_2}V_{\beta}(x_1,x_2), \tag{2.3.17}$$

and any stationary policy that realizes the minimum in (2.3.17) is optimal for the discounted cost criterion (Theorem 4.1.4 in [70]). Moreover, because costs are nonnegative and the control set is finite, the following proposition follows directly from Proposition 4.3.1 in [70].

Proposition 2.3. *For* $0<\beta<1$,

$$V_{\beta}(x_1,x_2)=\lim_{n\to\infty}V_{n,\beta}(x_1,x_2).$$

In the remaining of the section we will obtain properties of optimal policies for the infinite horizon problem by first deriving properties of $V_{n,\beta}(x_1,x_2)$ and then using Proposition 2.3 to get similar properties for $V_{\beta}(x_1,x_2)$. The following lemma establishes the monotonicity of the value function with respect to the number of jobs in the system.

Lemma 2.8. $V_{n,\beta}(x_1,x_2)$ *is nondecreasing in* x_1 *and* x_2.

Letting $g_{n,\beta}(x_1,x_2) = V_{n,\beta}(x_1,(x_2-1)^+) - V_{n,\beta}(x_1,x_2)$ and $g_\beta(x_1,x_2) = \lim_{n\to\infty} g_{n,\beta}(x_1,x_2)$, we get from (2.3.16) that the incentive to allocate additional service rate ρ to Station 2 is equal to $T_{\rho_1,\rho_2}V_\beta(x_1,x_2) - T_{\rho_1,\rho_2+\rho}V_\beta(x_1,x_2) = -\rho g_\beta(x_1,x_2) \geq 0$ by Lemma 2.8. Therefore, Proposition 2.1 holds with W_{ρ_1,ρ_2} replaced with $T_{\rho_1,\rho_2}V_\beta$, which means that the dedicated server in Station 2 is always kept busy under the optimal policy. The corresponding incentive for Station 1 is equal to $T_{\rho_1,\rho_2}V_\beta(x_1,x_2) - T_{\rho_1+\rho,\rho_2}V_\beta(x_1,x_2) = \rho f_\beta(x_1,x_2)$, where $f_\beta(x_1,x_2) = \lim_{n\to\infty} f_{n,\beta}(x_1,x_2)$ and $f_{n,\beta}(x_1,x_2) = V_{n,\beta}(x_1,x_2) - V_{n,\beta}(x_1-1,x_2+1)$, $x_1 \geq 1$. Therefore, Proposition 2.2 also holds with W_{ρ_1,ρ_2} and f replaced with $T_{\rho_1,\rho_2}V_\beta$ and f_β, respectively. The following is the counterpart of Lemma 2.3.

Lemma 2.9. *Let $h_1 \geq h_2$. Then, $f_{n,\beta}(x_1,x_2) \geq 0$ for all $x_1 \geq 1$, $x_2 \geq 0$.*

As a consequence, non-idling policies are optimal when $h_1 \geq h_2$. In that case, to determine where the flexible server should be assigned we use decision functions $d_{n,\beta}$, $\tilde{d}_{n,\beta}$, $\hat{d}_{n,\beta}$, and $\bar{d}_{n,\beta}$, defined in terms of $V_{n,\beta}$ in the same way it was done for clearing systems (Equations (2.3.11)-(2.3.14)), with d_β, $\tilde{d}_\beta$, $\hat{d}_\beta$, and $\bar{d}_\beta$ being the corresponding limits as $n \to \infty$.

For the general model with dedicated servers in both stations we have derived asymptotic properties of the optimal policy for a large number of jobs in the downstream station. They are given in the following theorem.

Theorem 2.9. *i) Assume $h_1 \geq h_2$ and $\mu_1(h_1-h_2) - \mu_2 h_2 < 0$. Then the optimal policy assigns the flexible server to Station 2 for $x_2 \geq Y_1+1$, where $Y_1 = \min\left\{x \middle| \beta^x < \frac{(\mu_1+\mu_2)h_2-\mu_1 h_1}{(\mu_1+\mu_2)h_2}\right\}$.*
ii) Assume $h_1 \geq h_2$ and $\mu_1(h_1-h_2) - \xi_2 h_2 < 0$. Then, if $\beta < \frac{(\mu_1+\xi_2)h_2-\mu_1 h_1}{(\mu_1+\xi_2)h_2}$, the optimal policy assigns the flexible server to Station 2 for all $x_2 \geq 1$.
iii) Assume $h_1 < h_2$ and define $Y_2 = \min\left\{x \middle| \beta^x < \frac{h_2-h_1}{h_2}\right\}$. Then, the optimal policy idles the dedicated server in Station 1 and assigns the flexible server to Station 2 for $x_2 \geq Y_2 - \mathbf{1}(Y_2 = 1)$.

Part (iii) is an extension of Theorem 4.2 in [64] where the model without a flexible server is studied. It is worth noting that when the discount factor is sufficiently small ($\beta < (h_2-h_1)/h_2$), the optimal policy never serves Station 1.

For systems with no dedicated server in Station 1 the following theorem extends results for fully collaborative servers ([88],[57]) to the case of partially collaborative servers. It is essentially a repetition of Theorem 2.3 for clearing systems without the part on the slope of the switching curve that determines the optimal policy.

Theorem 2.10. *Assume $h_1 \geq h_2$, $\nu_1 = 0$, $\nu_2 \geq \mu_2$, and partially collaborative servers in Station 2. Then*
i) When $\mu_1(h_1-h_2) < \mu_2 h_2$, for each $x_1 \geq 1$, there exists an integer $t(x_1) \geq 1$ such that the optimal policy assigns the flexible server to Station 2 (resp. 1) when $x_2 \geq t(x_1)$ (resp. $x_2 < t(x_1)$).
ii) When $\mu_1(h_1-h_2) \geq \mu_2 h_2$, the optimal policy assigns the flexible server to Station 1 for all $x_1 \geq 1$.

For systems with no dedicated server in Station 2 we have obtained a result analogous to part (i) of Theorem 2.8. Note that we have only managed to prove this result for fully collaborative servers.

Theorem 2.11. *Assume $h_1 < h_2$, $\nu_2 = 0$, $\mu_1 \geq \mu_2$, and fully collaborative servers in Station 1. Then the optimal policy assigns the flexible server to Station 2 for all $x_2 \geq 1$.*

Having shown that the flexible server should be assigned to Station 2 when there are jobs there, our model becomes similar to the model of [64] where there is one dedicated server in each station. The difference between the two models is that in our case the flexible server can move to Station 1 when there are no jobs in Station 2. In [64] it was shown that the dedicated server in Station 1 should idle if the number of jobs in Station 2 exceeds a threshold (as in part (ii) of Theorem 2.8). We make the conjecture that this is true for our model as well.

2.3.2.2 Average cost

Let e_i, $i = 1, 2$, be the policy that never idles the dedicated servers and assigns the flexible server to Station i. Then, the system is stable under e_1 and e_2 if $\lambda < \min\{\nu_1+\mu_1, \nu_2\}$ and $\lambda < \min\{\nu_1, \nu_2+\mu_2\}$, respectively. Therefore, if $\lambda < \Lambda = \max\{\min\{\nu_1 + \mu_1, \nu_2\}, \min\{\nu_1, \nu_2 + \mu_2\}\}$, there exists a stationary policy e (e_1 or e_2) that induces an irreducible and positive recurrent Markov chain on the state space and for all x_1, x_2 satisfies $J^e = J^e(x_1, x_2) < \infty$, which also implies that the set $\{(x_1, x_2) : h_1x_1 + h_2x_2 < J^e\}$ is finite. Because of the aforementioned conditions, Proposition 4.3 in [69] is valid, according to which a set of assumptions hold that lead to Theorem 7.2.3 in [70], which in turn results in Proposition 2.4 that follows. First, for some fixed state (z_1, z_2) with $V_\beta(z_1, z_2) < \infty$,[1] we define w to be a limit function if there exists a sequence $\beta_l \to 1$ for which

$$\lim_{l\to\infty} [V_{\beta_l}(x_1, x_2) - V_{\beta_l}(z_1, z_2)] = w(x_1, x_2). \tag{2.3.18}$$

Proposition 2.4. *Let $\lambda < \Lambda$. Then, there exist a finite constant J and a limit function w satisfying*

$$J + w(x_1, x_2) \geq h_1x_1 + h_2x_2 + \min_{(\rho_1,\rho_2)\in A(x_1,x_2)} T_{\rho_1,\rho_2}w(x_1, x_2).$$

If π is a stationary policy realizing the minimum, π is average cost optimal with average cost J.

It is clear from Proposition 2.4 and (2.3.18) that any properties of optimal policies under the discounted cost criterion that have been shown to hold for every $\beta < 1$ in the previous section are also valid for the average cost problem for $\lambda < \Lambda$. Specifically, idling the dedicated server in Station 2 is not optimal, non-idling in general is optimal for $h_1 \geq h_2$ and Theorem 2.11 holds for $\lambda < \min\{\nu_1, \mu_2\}$.

For systems with no dedicated server in Station 1 we have not been able to prove the existence of a switching curve as we did for the discounted cost problem in Theorem 2.10.

[1] $V_\beta(z_1, z_2) < \infty$ for every z_1, z_2 (see, e.g., Lemma 2.1 of [64]).

Proposition 2.4 and (2.3.18) indicate that the decision function is non-increasing with the number of jobs in Station 2, but to prove that it becomes negative would require a result analogous to part (i) of Theorem 2.9, but $Y_1 \to \infty$ as $\beta \to 1$. Therefore, we have the following characterization of the optimal policy.

Theorem 2.12. *Assume $h_1 \geq h_2$, $\nu_1 = 0$, $\nu_2 \geq \mu_2$, $\lambda < \min\{\mu_1, \nu_2\}$, and partially collaborative servers in Station 2. Then*
i) When $\mu_1(h_1 - h_2) < \mu_2 h_2$, if the optimal policy assigns the flexible server to Station 2 at state (x_1, x_2), it does so at state $(x_1, x_2 + 1)$ as well.
ii) When $\mu_1(h_1 - h_2) \geq \mu_2 h_2$, the optimal policy always assigns the flexible server to Station 1.

2.4 Non-preemptive service discipline

We assume that the flexible server is slower than each dedicated server and server collaboration is not allowed, that is, $\nu_i > \mu_i$ and $\xi_i = 0$, $i = 1, 2$. Therefore, we have a two-stage extension of the slow server problem. Compared to the model with preemptions, this problem is more difficult to analyze because the state space has to include the servers' state in addition to the number of jobs in each station. To the best of our knowledge, this is the first attempt to study the slow server problem in the context of tandem queueing systems. We characterize the optimal policy for the fast servers and provide conjectures for the slow server that are supported by extensive numerical experiments.

2.4.1 Clearing systems

We formulate the problem as a Markov decision process with state space (x_1, i_1, j, x_2, i_2), where x_1, x_2 is the number of jobs waiting in Station 1 and 2 respectively, i_1, i_2 is the state of the dedicated server in Station 1 and 2 (0 when idle, 1 when busy), and j is the state of the slow server (0 when idle, 1 when working in Station 1, 2 when working in Station 2). Instead of the continuous time problem we consider an equivalent discrete time problem obtained by uniformization, where without loss of generality we assume $\nu_1 + \nu_2 + \mu_1 + \mu_2 = 1$. We denote by $V(x_1, i_1, j, x_2, i_2)$ the minimum expected cost starting from state (x_1, i_1, j, x_2, i_2) with server allocations pending. Then, the following optimality equation is satisfied.

$$V(x_1, i_1, j, x_2, i_2) = \min_{\alpha_1, \alpha_2, \alpha} W(x_1', \alpha_1, \alpha, x_2', \alpha_2), \tag{2.4.1}$$

where $\alpha_1, \alpha_2, \alpha$ are possible allocations for the dedicated servers and the flexible server (taking values similarly to i_1, i_2, j), $x_k' = x_k - \alpha_k(1 - i_k) - 1(j = 0, \alpha = k)$, $k = 1, 2$, is the number of jobs waiting as a result of actions $\alpha_1, \alpha_2, \alpha$, and $W(x_1, i_1, j, x_2, i_2)$ is the minimum expected cost starting from state (x_1, i_1, j, x_2, i_2) after decisions have been made. For $j = 0, 1$ we have

$$\begin{aligned} W(x_1, i_1, j, x_2, i_2) &= h_1(x_1 + i_1 + j) + h_2(x_2 + i_2) + \nu_1 V(x_1, 0, j, x_2 + i_1, i_2) \\ &\quad + \nu_2 V(x_1, i_1, j, x_2, 0) + \mu_1 V(x_1, i_1, 0, x_2 + j, i_2) + \mu_2 V(x_1, i_1, j, x_2, i_2), \end{aligned} \tag{2.4.2}$$

for $j = 2$,

$$\begin{aligned} W(x_1, i_1, 2, x_2, i_2) &= h_1(x_1 + i_1) + h_2(x_2 + i_2 + 1) + \nu_1 V(x_1, 0, 2, x_2 + i_1, i_2) \\ &+\nu_2 V(x_1, i_1, 2, x_2, 0) + \mu_1 V(x_1, i_1, 2, x_2, i_2) + \mu_2 V(x_1, i_1, 0, x_2, i_2), \end{aligned} \quad (2.4.3)$$

and

$$V(0, 0, 0, 0, 0, 0) = 0. \quad (2.4.4)$$

We use (2.4.1)-(2.4.4) in the following section to characterize the optimal policy regarding the fast servers.

2.4.1.1 Optimal allocation of fast servers

First we show in Theorem 2.13 that there is no incentive to idle the dedicated server of Station 2 because we need to push jobs out of the system as fast as possible. The proof of the theorem requires some of the properties of the value function given in the following lemma.

Lemma 2.10. *i)* $V(x_1, i_1, j, x_2, i_2) \leq V(x_1, i_1, j, x_2 + 1, i_2)$,
ii) $V(x_1, i_1, j, x_2, 0) \leq V(x_1, i_1, j, x_2, 1)$,
iii) $V(x_1, i_1, 0, x_2, i_2) \leq V(x_1, i_1, 2, x_2, i_2)$,
iv) $V(x_1, i_1, j, x_2 - 1, 1) \leq V(x_1, i_1, j, x_2, 0)$,
v) $V(x_1, i_1, 0, x_2, 1) \leq V(x_1, i_1, 2, x_2, 0)$.

Proof. See Appendix B. □

Theorem 2.13. *The optimal policy does not idle the dedicated server at Station 2.*

Proof. It suffices to prove the following properties.

$$W(x_1, i_1, j, x_2 - 1, 1) \leq W(x_1, i_1, j, x_2, 0), \quad x_2 \geq 1, \quad (2.4.5)$$

$$W(x_1, i_1, 0, 0, 1) \leq W(x_1, i_1, 2, 0, 0). \quad (2.4.6)$$

Eq. (2.4.5) implies that any policy that idles the dedicated server of Station 2 can be improved by a policy that assigns a job to this server and keeps the allocation of the other servers unchanged. According to (2.4.6), if there is one job at Station 2 and both the dedicated server of Station 2 and the slow server are available, it is better to assign it the fast rather than the slow server. Using (2.4.2) and (2.4.3) we get

$$\begin{aligned} &W(x_1, i_1, j, x_2 - 1, 1) - W(x_1, i_1, j, x_2, 0) \\ &= \nu_1[V(x_1, 0, j, x_2 - 1 + i_1, 1) - V(x_1, 0, j, x_2 + i_1, 0)] \\ &+\nu_2[V(x_1, i_1, j, x_2 - 1, 0) - V(x_1, i_1, j, x_2, 0)] \\ &+\mu_1[V(x_1, i_1, j \cdot \mathbf{1}(j \neq 1), x_2 - 1 + \mathbf{1}(j = 1), 1) \\ &-V(x_1, i_1, j \cdot \mathbf{1}(j \neq 1), x_2 + \mathbf{1}(j = 1), 0)] \\ &+\mu_2[V(x_1, i_1, \mathbf{1}(j = 1), x_2 - 1, 1) - V(x_1, i_1, \mathbf{1}(j = 1), x_2, 0)], \end{aligned} \quad (2.4.7)$$

and

$$\begin{aligned} & W(x_1,i_1,0,0,1) - W(x_1,i_1,2,0,0) \\ & \quad = \nu_1[V(x_1,0,0,i_1,1) - V(x_1,0,2,i_1,0)] \\ & \quad +\nu_2[V(x_1,i_1,0,0,0) - V(x_1,i_1,2,0,0)] \\ & \quad +\mu_1[V(x_1,i_1,0,0,1) - V(x_1,i_1,2,0,0)] \\ & \quad +\mu_2[V(x_1,i_1,0,0,1) - V(x_1,i_1,0,0,0)]. \end{aligned} \tag{2.4.8}$$

The righthand side of (2.4.7) is non-positive because of part (i) of Lemma 2.10 (term multiplying ν_2) and part (iv) of Lemma 2.10 (remaining terms), proving (2.4.5). The term multiplying μ_2 in (2.4.8) is nonnegative because of part (ii) of Lemma 2.10. Then, taking into account that $\mu_2 < \nu_2$, we get

$$\begin{aligned} & W(x_1,i_1,0,0,1) - W(x_1,i_1,2,0,0) \\ & \quad \le \nu_1[V(x_1,0,0,i_1,1) - V(x_1,0,2,i_1,0)] \\ & \quad +(\nu_2+\mu_1)[V(x_1,i_1,0,0,1) - V(x_1,i_1,2,0,0)] \le 0, \end{aligned}$$

because of part (v) of Lemma 2.10, proving (2.4.6). □

Because of Theorem 2.13 the state of the system can be written more compactly as (x_1, i_1, j, x_2), where x_2 is the number of jobs in station 2 including the one assigned to the fast server. Then, the optimality equation takes the form

$$V(x_1,i_1,j,x_2) = \min_{\alpha_1,\alpha} W(x_1', \alpha_1, \alpha, x_2'), \tag{2.4.9}$$

where $x_1' = x_1 - \alpha_1(1-i_1) - \mathbf{1}(j=0,\alpha=1)$, $x_2' = x_2 - \mathbf{1}(j=0,\alpha=2)$, for $j = 0,1$

$$\begin{aligned} W(x_1,i_1,j,x_2) &= h_1(x_1+i_1+j) + h_2x_2 + \nu_1 V(x_1,0,j,x_2+i_1) \\ & +\nu_2 V(x_1,i_1,j,(x_2-1)^+) + \mu_1 V(x_1,i_1,0,x_2+j) + \mu_2 V(x_1,i_1,j,x_2), \end{aligned} \tag{2.4.10}$$

for $j = 2$,

$$\begin{aligned} W(x_1,i_1,2,x_2) &= h_1(x_1+i_1) + h_2(x_2+1) + \nu_1 V(x_1,0,2,x_2+i_1) \\ & +\nu_2 V(x_1,i_1,2,(x_2-1)^+) + \mu_1 V(x_1,i_1,2,x_2) + \mu_2 V(x_1,i_1,0,x_2), \end{aligned} \tag{2.4.11}$$

and

$$V(0,0,0,0) = 0. \tag{2.4.12}$$

As a consequence of the new state space definition, parts (i) and (ii) of Lemma 2.10 can be combined to give

$$V(x_1,i_1,j,x_2) \le V(x_1,i_1,j,x_2+1). \tag{2.4.13}$$

When it is not cheaper to have jobs in Station 1 compared to Station 2, it is reasonable that at least the fast server in Station 1 should never idle because there is no incentive to keep jobs upstream. This is shown in Theorem 2.14, whose proof requires properties of the value function given in the following lemma.

Lemma 2.11. *Let $h_1 \geq h_2$. Then,*
i) $V(x_1, i_1, j, x_2) \leq V(x_1 + 1, i_1, j, x_2)$,
ii) $V(x_1, 0, j, x_2) \leq V(x_1, 1, j, x_2)$,
iii) $V(x_1, i_1, 0, x_2) \leq V(x_1, i_1, 1, x_2)$,
iv) $V(x_1 - 1, i_1, j, x_2 + 1) \leq V(x_1, i_1, j, x_2), \quad x_1 \geq 1$,
v) $V(x_1, 0, j, x_2 + 1) \leq V(x_1, 1, j, x_2)$,
vi) $V(x_1, i_1, 0, x_2 + 1) \leq V(x_1, i_1, 1, x_2)$,
vii) $V(x_1 - 1, 1, j, x_2) \leq V(x_1, 0, j, x_2), \quad x_1 \geq 1$,
viii) $V(x_1, 1, 0, x_2) \leq V(x_1, 0, 1, x_2)$.

Proof. See Appendix B. □

Theorem 2.14. *If $h_1 \geq h_2$, the optimal policy does not idle the dedicated server at Station 1.*

Proof. It suffices to prove the following properties.

$$W(x_1 - 1, 1, j, x_2) \leq W(x_1, 0, j, x_2), \quad x_2 \geq 1, \tag{2.4.14}$$

$$W(0, 1, 0, x_2) \leq W(0, 0, 1, x_2). \tag{2.4.15}$$

Eq. (2.4.14) implies that any policy that idles the dedicated server of Station 1 can be improved by a policy that assigns a job to this server and keeps the allocation of the other servers unchanged. According to (2.4.15), if there is one job at Station 1 and both the dedicated server of Station 1 and the slow server are available, it is better to assign it to the fast rather than the slow server. Using (2.4.10) and (2.4.11) we get

$$\begin{aligned}
&W(x_1 - 1, 1, j, x_2) - W(x_1, 0, j, x_2) \\
&\quad = \nu_1[V(x_1 - 1, 0, j, x_2 + 1) - V(x_1, 0, j, x_2)] \\
&\quad +\nu_2[V(x_1 - 1, 1, j, (x_2 - 1)^+) - V(x_1, 0, j, (x_2 - 1)^+)] \\
&\quad +\mu_1[V(x_1 - 1, 1, j \cdot \mathbf{1}(j \neq 1), x_2 + \mathbf{1}(j = 1)) \\
&\quad -V(x_1, 0, j \cdot \mathbf{1}(j \neq 1), x_2 + \mathbf{1}(j = 1))] \\
&\quad +\mu_2[V(x_1 - 1, 1, \mathbf{1}(j = 1), x_2) - V(x_1, 0, \mathbf{1}(j = 1), x_2)],
\end{aligned} \tag{2.4.16}$$

and

$$\begin{aligned}
&W(0, 1, 0, x_2) - W(0, 0, 1, x_2) \\
&\quad = \nu_1[V(0, 0, 0, x_2 + 1) - V(0, 0, 1, x_2)] \\
&\quad +\nu_2[V(0, 1, 0, (x_2 - 1)^+) - V(0, 0, 1, (x_2 - 1)^+)] \\
&\quad +\mu_1[V(0, 1, 0, x_2) - V(0, 0, 0, x_2 + 1)] \\
&\quad +\mu_2[V(0, 1, 0, x_2) - V(0, 0, 1, x_2)].
\end{aligned} \tag{2.4.17}$$

The righthand side of (2.4.16) is non-positive because of part (iv) of Lemma 2.11 (term multiplying ν_1) and part (vii) of Lemma 2.11 (remaining terms), proving (2.4.14). The term multiplying μ_1 in (2.4.17) is nonnegative because of part (v) of Lemma 2.11. Then, taking into account that $\mu_1 < \nu_1$, we get

$$\begin{aligned} &W(0,1,0,x_2) - W(0,0,1,x_2) \\ &\quad \leq \nu_2[V(0,1,0,(x_2-1)^+) - V(0,0,1,(x_2-1)^+)] \\ &\quad +(\nu_1+\mu_2)[V(0,1,0,x_2) - V(0,0,1,x_2)] \leq 0, \end{aligned}$$

because of part (viii) of Lemma 2.11, proving (2.4.15). □

We have not been able to obtain qualitative properties of the optimal allocation strategy for the slow server and the dedicated server of Station 1 when $h_1 < h_2$. In the next section we present numerical results that illustrate the structure of the optimal policy.

2.4.1.2 Numerical investigation

When $h_1 \geq h_2$ the decision to be made is how to use the slow server. Our numerical investigation led us to conjecture that the optimal policy has the following two-threshold structure. For a given number of jobs in Station 1, the flexible server is assigned to Station 1, is kept idle, and assigned to Station 2 when the number of jobs in Station 2 is below the lower threshold, between the two thresholds, and above the upper threshold, respectively. This is illustrated in the following example.

Example 2.3. *Let $\nu_1 = 2$, $\nu_2 = 4$, $\mu_1 = 0.2$, $\mu_2 = 0.1$, $h_1 = 2$, and $h_2 = 1.5$. Then, the optimal allocation of the slow server is given in Figure 2.3.*

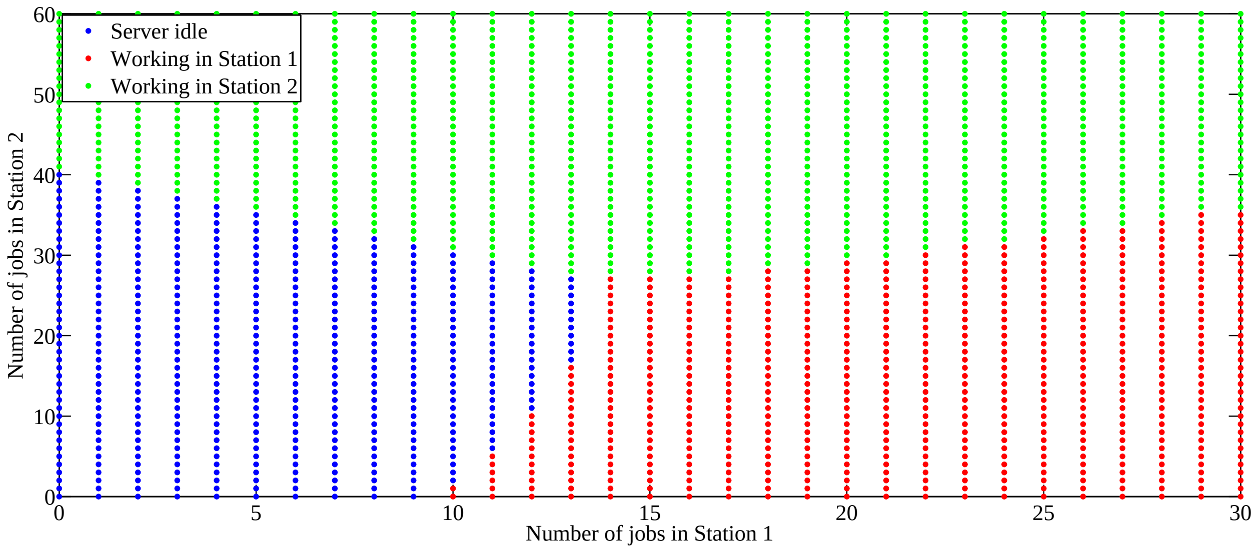

Figure 2.3: Slow server allocation

When $h_1 < h_2$ it may be optimal to idle the dedicated server of Station 1. Therefore, there are at most 6 possible allocations at a decision instant. Regarding the slow server allocation, our conjecture on the aforementioned two-threshold structure of the optimal policy was verified by our numerical results, which also indicated that the threshold values that may also depend on the state of the dedicated server of Station 1. Furthermore, when both the dedicated server of Station 1 and the flexible server are available at a decision instant, our numerical results indicated that it is not optimal to idle the fast server and assign the slow server to Station 1.

A special case of interest is when the slow server is constrained to work in one of the stations. We first consider the case when it is constrained to work downstream. In this case Station 2 is fed by Station 1, so the problem of the optimal use of the slow server resembles the classical slow server problem with arrivals, leading to the conjecture that the slow server is used if the number of jobs in Station 2 exceeds a certain threshold. We further conjecture that this threshold value in non-increasing with the number of jobs in Station 1 because the arrival rate seen by Station 2 is increasing in the number of jobs upstream. In addition, we believe that for a sufficiently large number of jobs in Station 1 and $\nu_1 < \nu_2 + \mu_2$ the problem is equivalent to the classical slow server problem with arrival rate equal to ν_1, the service rate of the dedicated server of Station 1. Our conjectures were verified by the combined use of numerical experiments (two-stage systems) and the algorithm provided by Lin and Kumar [51] for computing the threshold determining the optimal policy under the average cost criterion for single stage systems with arrivals. The following is a relevant example with $h_1 \geq h_2$.

Example 2.4. *Let* $\nu_1 = 0.82$, $\nu_2 = 0.91$, $\mu_2 = 0.12$, $h_1 = 3$, *and* $h_2 = 1$. *Then, the optimal use of the slow server is determined by the switching curve given in Figure 2.4.*

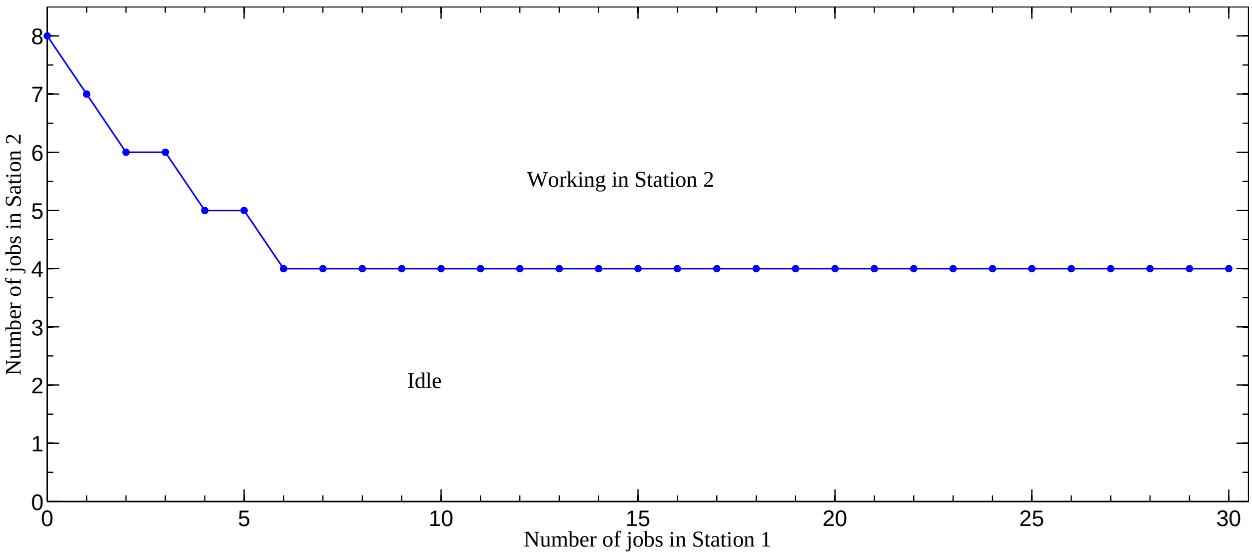

Figure 2.4: Switching points for the slow server

When $h_1 < h_2$ we have a variation of the two-stage model studied by Rosberg et al. [64] which differs from ours in that there are arrivals and no slow server. For their model they showed that it is optimal to idle the dedicated server of Station 1 when the number of jobs in Station 2 exceeds a threshold which is non-decreasing in the number of jobs in Station

1. Our numerical results were in agreement with this property of the optimal policy. The following example differs from the previous one in the holding cost rates. We have kept the same service rates in order to show that the asymptotic behavior of the switching curve for the slow server depends only on these rates.

Example 2.5. *Let* $\nu_1 = 0.82$*,* $\nu_2 = 0.91$*,* $\mu_2 = 0.12$*,* $h_1 = 2$*, and* $h_2 = 3$*. Then, the switching curves that determine whether the dedicated server of Station 1 and the slow server should be used or not are given in Figures 2.5 and 2.6.*

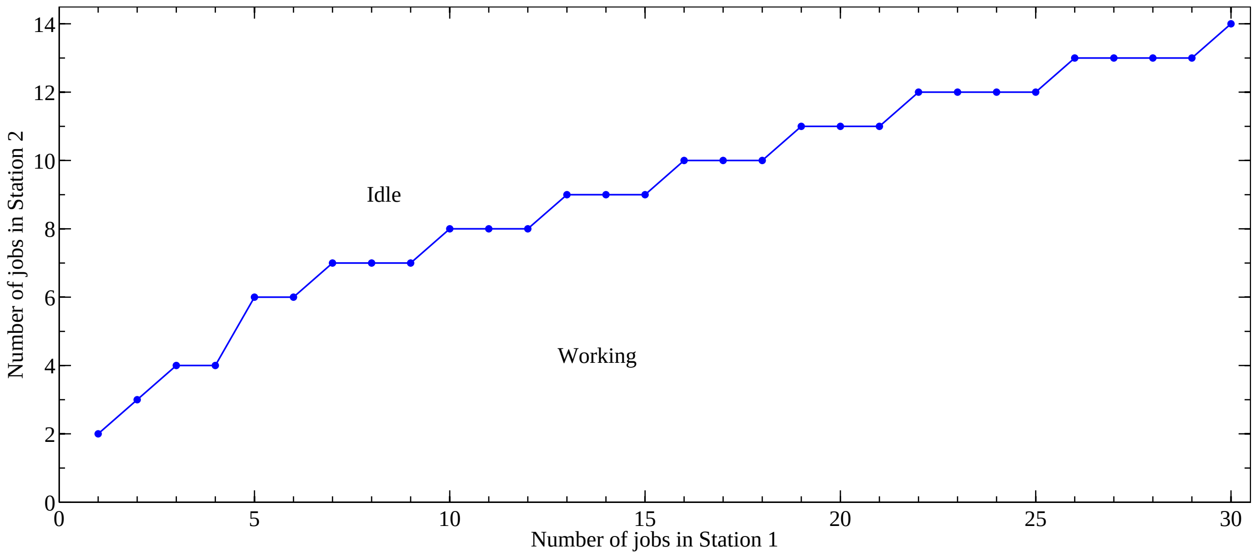

Figure 2.5: Shutdown points for the fast server in Station 1

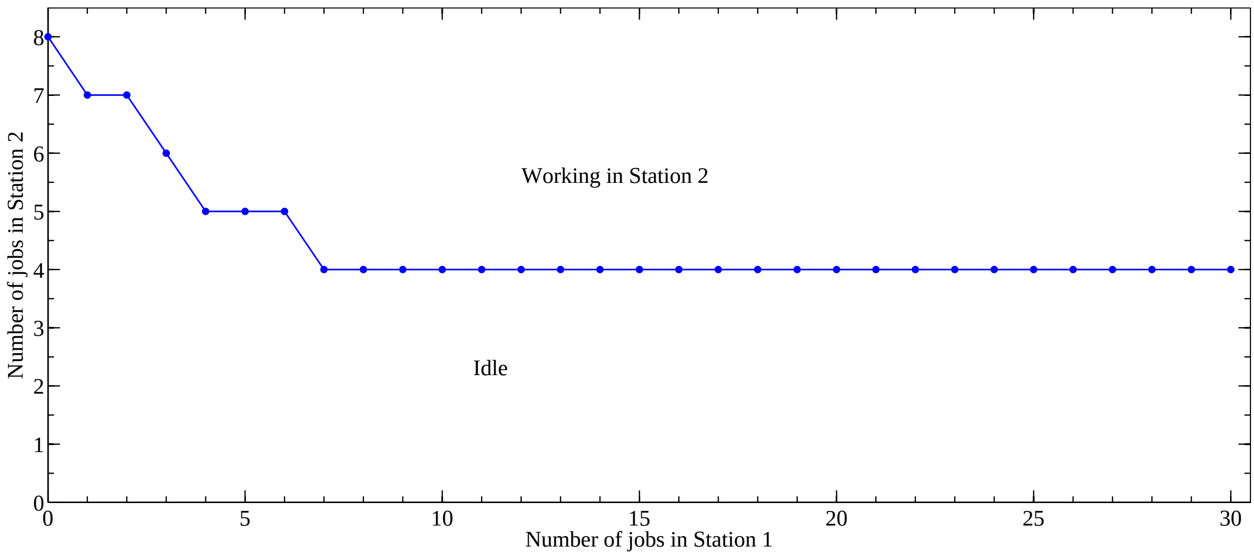

Figure 2.6: Switching points for the slow server

We now consider the case when the slow server is constrained to work in the upstream station. When $h_1 \geq h_2$, in which case the decision to be made is whether the slow server should be used or not, our numerical work indicated that the optimal policy is of threshold type. For a given number of jobs in Station 2, the slow server is used if the number of jobs in Station 1 exceeds a threshold. It is interesting that the switching curve defined by the threshold values may not be monotonic with respect to the number of jobs in Station 2, as shown in the next example. Note that a similar behavior is observed in Example 2.3.

Example 2.6. *Let* $\nu_1 = 0.11$, $\nu_2 = 0.37$, $\mu_1 = 0.02$, $h_1 = 3$, *and* $h_2 = 2.5$. *Then, the optimal use of the slow server is depicted in Figure 2.7.*

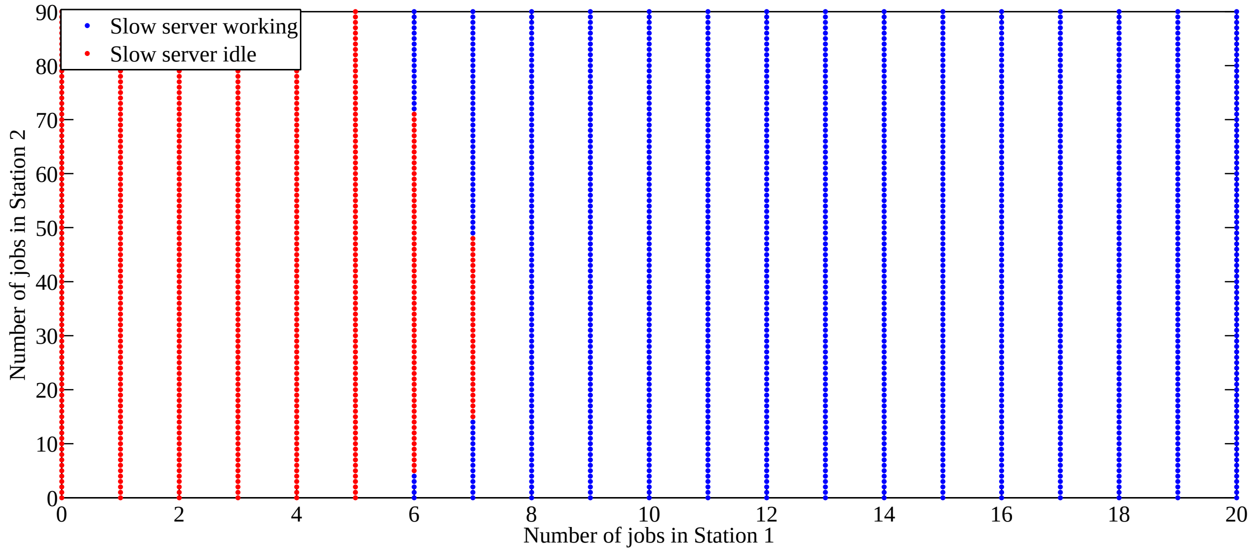

Figure 2.7: Slow server allocation

When $h_1 < h_2$ our numerical work focused on decision instants at which both the dedicated server of Station 1 and the slow server are available, so that a simultaneous allocation for both servers needs to be made. We reached the conclusion that the optimal strategy is characterized by two thresholds. For a given number of jobs in Station 1, both servers are assigned to jobs, only the dedicated server is assigned to a job, and both servers idle when the number of jobs in Station 2 is below the lower threshold, between the two thresholds, and above the upper threshold, respectively. This is illustrated in the following example.

Example 2.7. *Let* $\nu_1 = 0.11$, $\nu_2 = 0.37$, $\mu_1 = 0.02$, $h_1 = 0.85$, *and* $h_2 = 2.5$. *Then, the optimal allocation of the two servers is given in Figure 2.8.*

2.4.2 Systems with arrivals

In this section we study the previous model with jobs arriving at the upstream station according to a Poisson process with rate λ. We consider the equivalent discrete time problem obtained by uniformization, where without loss of generality we assume $\lambda + \nu_1 + \nu_2 + \mu_1 + \mu_2 = 1$. We denote by $V_n(x_1, i_1, j, x_2, i_2)$ the minimum n-step expected cost starting from state

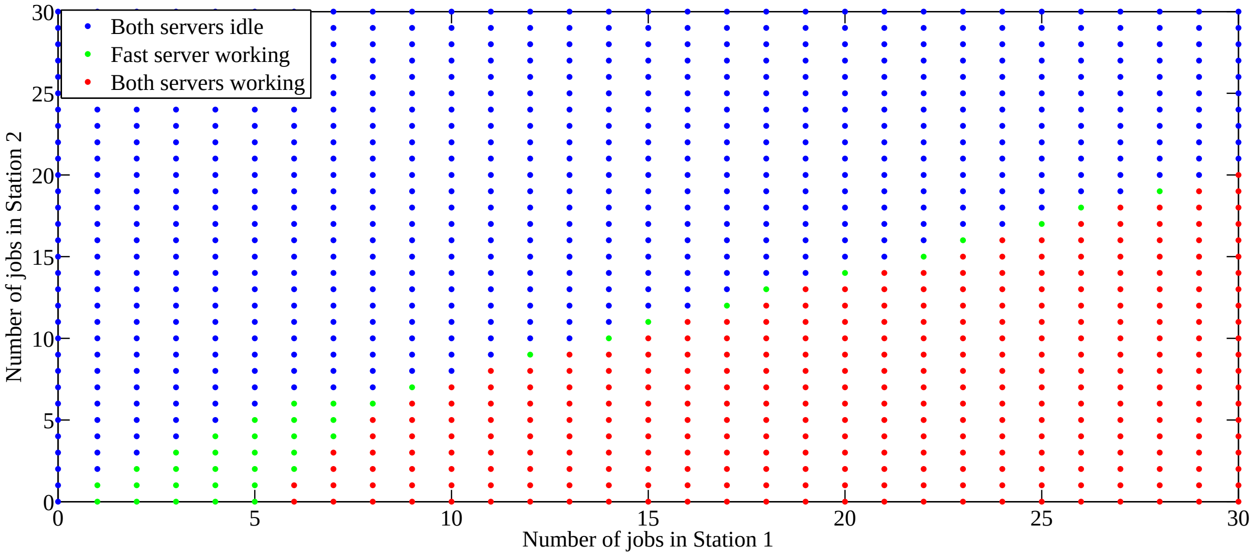

Figure 2.8: Slow server allocation

(x_1, i_1, j, x_2, i_2) before decisions have been made. Then, the following optimality equation is satisfied.

$$V_n(x_1, i_1, j, x_2, i_2) = \min_{\alpha_1,\alpha_2,\alpha} W_n(x_1', \alpha_1, \alpha, x_2', \alpha_2), \tag{2.4.18}$$

where $\alpha_1, \alpha_2, \alpha$ are possible allocations for the dedicated servers and the flexible server, x_k', $k = 1, 2$, is the number of jobs waiting as a result of actions $\alpha_1, \alpha_2, \alpha$, and $W_n(x_1, i_1, j, x_2, i_2)$ is the minimum n-step expected cost starting from state (x_1, i_1, j, x_2, i_2) after server allocations have been finalized. With β being the discount factor, for $j = 0, 1$ we have

$$\begin{aligned} W_n(x_1, i_1, j, x_2, i_2) &= h_1(x_1 + i_1 + j) + h_2(x_2 + i_2) + \beta\,[\lambda V_{n-1}(x_1 + 1, i_1, j, x_2, i_2) \\ &\quad + \nu_1 V_{n-1}(x_1, 0, j, x_2 + i_1, i_2) + \nu_2 V_{n-1}(x_1, i_1, j, x_2, 0) \\ &\quad + \mu_1 V_{n-1}(x_1, i_1, 0, x_2 + j, i_2) + \mu_2 V_{n-1}(x_1, i_1, j, x_2, i_2)]\,, \end{aligned} \tag{2.4.19}$$

for $j = 2$,

$$\begin{aligned} W_n(x_1, i_1, 2, x_2, i_2) &= h_1(x_1 + i_1) + h_2(x_2 + i_2 + 1) + \beta\,[\lambda V_{n-1}(x_1 + 1, i_1, 2, x_2, i_2) \\ &\quad + \nu_1 V_{n-1}(x_1, 0, 2, x_2 + i_1, i_2) + \nu_2 V_{n-1}(x_1, i_1, 2, x_2, 0) \\ &\quad + \mu_1 V_{n-1}(x_1, i_1, 2, x_2, i_2) + \mu_2 V_{n-1}(x_1, i_1, 0, x_2, i_2)]\,, \end{aligned} \tag{2.4.20}$$

and

$$V_0(x_1, i_1, j, x_2, i_2) = 0. \tag{2.4.21}$$

We show that Theorems 2.13 and 2.14 hold for the finite horizon expected discounted cost problem, defined by (2.4.18)-(2.4.21). Then, using the analysis for systems operating under the preemptive service discipline (Section 2.3.2), we can show that Theorems 2.13 and 2.14 hold for the infinite horizon expected discounted cost problem and assuming stable systems, for the expected average cost problem as well.

To show that the optimal policy never idles the dedicated server of Station 2 (Theorem 2.13), it suffices to prove the following properties for W_n.

$$\begin{aligned} W_n(x_1, i_1, j, x_2 - 1, 1) &\leq W_n(x_1, i_1, j, x_2, 0), \quad x_2 \geq 1, && (2.4.22) \\ W_n(x_1, i_1, 0, 0, 1) &\leq W_n(x_1, i_1, 2, 0, 0). && (2.4.23) \end{aligned}$$

For this purpose we prove the following group of properties for both V_n and W_n (see Appendix B).

$$\begin{aligned} &\text{(P1)}: && f(x_1, i_1, j, x_2, i_2) \leq f(x_1, i_1, j, x_2 + 1, i_2), \\ &\text{(P2)}: && f(x_1, i_1, j, x_2, 0) \leq f(x_1, i_1, j, x_2, 1), \\ &\text{(P3)}: && f(x_1, i_1, 0, x_2, i_2) \leq f(x_1, i_1, 2, x_2, i_2), \\ &\text{(P4)}: && f(x_1, i_1, j, x_2 - 1, 1) \leq f(x_1, i_1, j, x_2, 0), \quad x_2 \geq 1, \\ &\text{(P5)}: && f(x_1, i_1, 0, x_2, 1) \leq f(x_1, i_1, 2, x_2, 0), \end{aligned}$$

where $f = V_n$ or W_n. The first three properties are monotonicity properties of the cost functions with respect to the number of jobs in Station 2, and (2.4.22) and (2.4.23) follow from (P4) and (P5) for W_n.

To prove that the optimal policy does not idle the dedicated server of Station 1 when $h_1 \geq h_2$ (Theorem 2.14), we first rewrite the optimality equation taking into account Theorem 2.13. We have

$$V_n(x_1, i_1, j, x_2) = \min_{\alpha_1, \alpha} W_n(x_1', \alpha_1, \alpha, x_2'), \qquad (2.4.24)$$

where for $j = 0, 1$

$$\begin{aligned} W_n(x_1, i_1, j, x_2) = {} & h_1(x_1 + i_1 + j) + h_2(x_2) + \beta\,[\lambda V_{n-1}(x_1 + 1, i_1, j, x_2) \\ & + \nu_1 V_{n-1}(x_1, 0, j, x_2 + i_1) + \nu_2 V_{n-1}(x_1, i_1, j, (x_2 - 1)^+) \\ & + \mu_1 V_{n-1}(x_1, i_1, 0, x_2 + j) + \mu_2 V_{n-1}(x_1, i_1, j, x_2)]\,, \qquad (2.4.25) \end{aligned}$$

for $j = 2$,

$$\begin{aligned} W_n(x_1, i_1, 2, x_2) = {} & h_1(x_1 + i_1) + h_2(x_2 + 1) + \beta\,[\lambda V_{n-1}(x_1 + 1, i_1, 2, x_2) \\ & + \nu_1 V_{n-1}(x_1, 0, 2, x_2 + i_1) + \nu_2 V_{n-1}(x_1, i_1, 2, (x_2 - 1)^+) \\ & + \mu_1 V_{n-1}(x_1, i_1, 2, x_2) + \mu_2 V_{n-1}(x_1, i_1, 0, x_2)]\,, \qquad (2.4.26) \end{aligned}$$

and

$$V_0(x_1, i_1, j, x_2, i_2) = 0. \qquad (2.4.27)$$

With the new state space definition properties (P1) and (P2) for V_n can be combined to give

$$V_n(x_1, i_1, j, x_2) \leq V_n(x_1, i_1, j, x_2 + 1). \qquad (2.4.28)$$

For Theorem 2.14 we need to show

$$\begin{aligned} W_n(x_1 - 1, 1, j, x_2) &\leq W_n(x_1, 0, j, x_2), \quad x_1 \geq 1, && (2.4.29) \\ W_n(0, 1, 0, x_2) &\leq W_n(0, 0, 1, x_2), && (2.4.30) \end{aligned}$$

whose proof requires the following group of properties for both V_n and W_n (see Appendix B for their proof).

$$
\begin{aligned}
&(\mathrm{Q1}): && f(x_1, i_1, j, x_2) \le f(x_1+1, i_1, j, x_2),\\
&(\mathrm{Q2}): && f(x_1, 0, j, x_2) \le f(x_1, 1, j, x_2),\\
&(\mathrm{Q3}): && f(x_1, i_1, 0, x_2) \le f(x_1, i_1, 1, x_2),\\
&(\mathrm{Q4}): && f(x_1-1, i_1, j, x_2+1) \le f(x_1, i_1, j, x_2), \quad x_1 \ge 1,\\
&(\mathrm{Q5}): && f(x_1, 0, j, x_2+1) \le f(x_1, 1, j, x_2),\\
&(\mathrm{Q6}): && f(x_1, i_1, 0, x_2+1) \le f(x_1, i_1, 1, x_2),\\
&(\mathrm{Q7}): && f(x_1-1, 1, j, x_2) \le f(x_1, 0, j, x_2), \quad x_1 \ge 1,\\
&(\mathrm{Q8}): && f(x_1, 1, 0, x_2) \le f(x_1, 0, 1, x_2),
\end{aligned}
$$

where $f = V_n$ or W_n. The first three properties are monotonicity properties of the cost functions with respect to the number of jobs in station 1. Properties (Q4)-(Q6) indicate that we have a cost reduction if a job is moved from station 1 to station 2. Finally, (2.4.29) and (2.4.30) follow from (Q7) and (Q8) for W_n.

Finally, with regard to the allocation of the slow server and the dedicated server of Station 1 when $h_1 < h_2$, we believe that our conjectures for clearing systems apply to systems with arrivals as well. Our belief was verified by numerical experiments for which we used the value iteration algorithm to determine the optimal server allocation for several problem instances.

2.5 Conclusions

We characterized optimal server allocations in two-stage tandem queueing systems with dedicated servers in each stage and one flexible server. In the class of preemptive policies most of our results were obtained assuming partial collaboration of servers working on the same job. Compared to systems with fully collaborative servers studied in the past, the assumption of partial collaboration only matters for states with one job in a station, because otherwise the optimal policy assigns the servers to different jobs. Although this is a minor difference, it made the problem more complex and its technical analysis much more difficult. We managed to show that under certain conditions on service rates and holding costs properties of the optimal policy for fully collaborative systems also hold under partial collaboration. For different cases we relied on numerical experiments. We found that the partial collaboration assumption may alter significantly the structure of the optimal policy resulting in policies that do not possess intuition-based properties that have been shown to hold for fully collaborative servers. Specifically, the optimal policy may not be transition monotone and in the case of no dedicated server in the second stage it may assign the flexible server to the upstream station even when we have larger cost savings from a service completion downstream.

We also considered a two-stage extension of the slow server problem (the flexible server is slower than the dedicated servers). Because of the increased complexity resulting from the fact that preemptions are not allowed in this model we were only able to obtain properties of the optimal allocation strategy for the fast servers. We showed that the fast server working downstream should not idle and the one upstream should not idle in case it is more expensive to keep jobs there. Regarding the flexible server, we had to resort to numerical experiments in order to get insights into the structure of the optimal allocation policy. As is the case with the single stage model, our numerical results indicated that the optimal policy is determined by threshold values related to the number of jobs in the system. The proof of this fact could be a topic for future research.

Chapter 3

Newsvendor Models with Unreliable and Backup Suppliers

3.1 Introduction

Motivated by situations where firms need to employ risk mitigation strategies in the presence of unreliable suppliers, we study newsvendor models with risky suppliers and a backup supplier whose capacity can be reserved at a cost. The following are the main types of supply risks that have been considered: i) *random disruptions*, where either the whole order or nothing is delivered, ii) *random yield*, where the delivered quantity is a random fraction of the quantity ordered, and iii) *random capacity*, where the delivered quantity cannot exceed the capacity of the supplier. A supplier's inability to fully satisfy his customers may be caused by unforeseen events, such as natural disasters and labor strikes, that may shut down production facilities (random disruptions), production of defective units (random yield), and unexpected downtime and limited availability of raw materials (random capacity). A variety of supply risk mitigation strategies have been proposed and analyzed in the literature, such as expanding the supplier base in order to spread the risk, investing to improve the supplier's reliability, and pricing the product after the supplier's uncertainty is resolved (see detailed literature review in the next section).

In this chapter we study models in which a reliable backup supplier is used for hedging against supply risks. When following this strategy, a firm places an order to the primary unreliable suppliers and buys the option to use the capacity of the backup supplier through a reservation contract. This option may be exercised later, either after the delivery from the primary supplier or after the demand becomes known as well. Such contracts have been used among others for the purchase of chemicals, commodity metals, semiconductors, and electric power (Kleindorfer and Wu [46]). The premium paid for reserving the capacity of a backup supplier (or similarly for establishing an in-house capability) reflects the costs for adjusting production to be able to respond to urgent requests (for example, rescheduling production, reallocating/adding resources, giving lower priority to other customers).

We consider models with primary suppliers that are subject to random yield or random capacity. In our base models we assume one primary unreliable supplier and a retailer facing demand for one product. When the supply risk is due to random yield and the retailer exercises the option to buy from the backup supplier after the demand is revealed, we also study extensions of the base model for two primary suppliers or two products with a common backup supplier. For all models we show that the profit function has a unique maximum, which enables us to derive properties for the optimal order and reservation quantities. Our findings include conditions for the use of the backup supplier to be profitable and conditions under which reservation is irrelevant, in the sense that the retailer eventually buys all of the reserved quantity. For the case of random capacity we also obtain interesting insights on the effect that various parameters of the model have on the optimal decisions of the retailer. Finally, our theoretical results are supplemented by conjectures based on numerical experiments.

3.2 Related literature

In this literature survey we refer to previous work that deals with various supply risk mitigation strategies used by retailers. There are also a lot of models that deal with supplier-retailer coordination in a game theoretic context. We do not include these models because they are not related to our work.

Starting with random disruption models, Parlar and Perry [61] determined numerically the optimal order policy for a model with known demand and two suppliers that are assumed to be available for an exponentially distributed amount of time. Gurler and Parlar [32] solved a generalized version of the previous problem with Erlang distributed availability times. Tomlin and Wang [80] studied a newsvendor model with multiple products and unreliable suppliers, for which they assessed the effect of the correlation of the product demands on the optimal use of the suppliers. Tomlin [78] analyzed a multiperiod problem with two suppliers, one subject to disruptions and one reliable, but more expensive. He showed that, depending on the supplier's availability and the nature of disruptions (frequent and short or rare and long), the optimal strategy can be one of the following three: carrying inventory, using both suppliers, and passive acceptance of the risk. In a similar problem, Dong and Tomlin [20] examined insurance as a risk mitigation strategy complementary to the use of a backup supplier. For a newsvendor model with two products, Tomlin [79] evaluated 12 risk mitigation strategies which are combinations of the following: using two suppliers, using a backup supplier when needed, and product substitution. For a model with deterministic demand and two suppliers, one reliable and one unreliable, Hu et al. [38] compared the following two strategies: place orders to both suppliers, and give incentives to the unreliable supplier to invest in capacity restoration. Finally, Hu and Kostamis [39] investigated a manufacturer's optimal ordering policy to unreliable and reliable suppliers when the market price is affected by the manufacturer's output.

We review random yield models next. Parlar and Wang [62] studied a single-period model with two unreliable suppliers for both deterministic and stochastic demand. They determined the optimal orders for deterministic demand and provided an approximate solution for stochastic demand. For a multiperiod problem with stochastic demand, Anupindi and Akella [13] showed that the optimal ordering policy at each period is determined by two critical values of starting inventory and it never places an order only to the more expensive supplier. Agrawal and Nahmias [2] were the first to study a model with more than two suppliers. Assuming a fixed order cost, deterministic demand, and normally distributed yields, they determined the optimal number of suppliers and the optimal order quantity in the case of identical suppliers. For the model without fixed order cost, stochastic demands, and a given number of suppliers, Yang et al. [91] determined the optimal order quantities using mathematical programming techniques. The supplier selection problem for various single-period and multiperiod models was studied by Federgruen and Yang [24],[25],[26],[27], who showed that in the absence of fixed order costs, suppliers are selected based on their cost and the orders to those selected depend on their reliabilities. Burke et al. [16] and Merzifonluoglou and Feng [53] showed that this insight may not hold in the case of minimum order commitments and fixed order costs, respectively. Wang et al. [84] compared two alternative risk mitigation strategies, using two suppliers and investing in the improvement of a single supplier's reliability. They showed that the first strategy should be preferred when the two suppliers differ in cost, while the second one should be preferred when they differ in reliability. Finally, Babich et al. [15] studied the effect of trade credit on supplier selection and optimal order quantities.

With regard to random capacity models, Wang et al. [84] also evaluated the aforementioned two strategies for this case and reached opposite conclusions; the first strategy should be preferred when the two suppliers differ in reliability, while the second one should be preferred when they differ in cost. Wu et al. [89] studied a periodic-review system with two suppliers, one reliable and one subject to random capacity, and they showed that the optimal policy is determined by a quota for the unreliable supplier and a base-stock level for the reliable one. Feng and Shi [28] considered a multiperiod model where in addition to supply diversification, they use dynamic pricing of the product to shape the demand and thus reduce the supply risk. Unlike most models in the literature, they showed that there exist situations for which the exclusive use of the most expensive supplier may be optimal. Another model with price dependent demand was studied by Li et al. [49] who also assumed that the suppliers' capacities are correlated. In the case of two suppliers they showed that supplier selection is based only on their cost, but this is not necessarily true for more than two suppliers. The same authors [50] assessed the relative effect of supply diversification and pricing for a similar model with two suppliers, one reliable and one subject to random capacity.

There is also research on models that do not belong to one of the three classes discussed so far (random disruptions, random yield, random capacity). In addition to their random yield model with two suppliers, Anupindi and Akella [13] studied a model where the whole order is either delivered in the current time period or in the next time period. They showed that the optimal policy is to place orders to both suppliers or only to the cheaper one. For the same model with discrete demand, Swaminathan and Shanthikumar [77] showed by examples that it may be optimal to order only from the more expensive supplier. Dada et al. [19] analyzed a newsvendor model with multiple unreliable suppliers under various types of reliability. Specifically, they considered models where the delivered quantity may depend on the order (with random yield as a special case) or may be independent of the order (with random capacity as a special case). They showed that cost takes precedence over reliability in the selection of suppliers and each selected supplier's order depends on his reliability.

The models that are analyzed in this thesis are mostly related to models that deal with contingency strategies that use a reliable backup supplier whose capacity has to be reserved in advance. Köle and Bakal [47] obtained the optimal strategy for a newsvendor model with a primary supplier that is subject to random disruptions, where the option to buy from the backup supplier is exercised with different levels of information regarding supply and demand. Huang and Xu [41] studied a similar model with two primary suppliers where the backup supplier is utilized after supply becomes known. On the other hand, Saghafian and Van Oyen [66] considered a model with one primary supplier and two products sharing the same backup supplier. For a single-period model with deterministic demand and a supplier that may experience both disruptions and random yield in his production, Chopra et al. [17] studied the effect of the two types of risk on the optimal policy. Schmitt and Snyder [68] considered the same model in a multiperiod setting and showed that the use of a single-period approximation may significantly alter the optimal order and reservation quantities. Similar models with stochastic demands were considered by Guo et al. [31], who investigated numerically the effect of the two types of supply risks on the optimal decisions of the firm, and Giri and Bardhan [30] who determined the optimal decisions when the retailer exercises the option to buy from the backup supplier after both supply and demand are revealed. Hou and Hu [37] studied a slightly different model where the retailer has also the option to cover the whole demand from the backup supplier at a higher price if the initially reserved quantity is not sufficient for that. For a single-period random yield model, He and Yang [34] determined the optimal order quantities of a retailer that has the option to cancel part of the order to the reliable supplier after the delivery from the risky supplier materializes. In the case of random capacity there exist models in the literature that are related to ours in the sense that they involve some form of capacity reservation. Jain and Silver [42] analyzed a single-period model with a supplier having a random capacity and a retailer having the option to ensure a minimum level of capacity by paying a premium. Serel [72] studied a similar model in a multi-product setting. Serel [71] also studied a multiperiod model with the buyer reserving the capacity of a primary supplier and also having the option to place an order with a backup unreliable supplier at the beginning of each period. Finally, single-period models with primary suppliers whose capacity could be reserved and a spot market with possibly insufficient capacity, where the buyer could place orders after the realization of demand, were studied by Jain and Hazra [43] (multiple primary suppliers) and Serel [73] (price dependent demand). Note that the models with more than one supplier ([71],[43],[73]) differ from ours in that the backup supplier is the one with random capacity.

3.3 Problem formulation and preliminaries

We consider a single-period model with a retailer facing random demand X and using two suppliers, a primary and a backup. The primary supplier is unreliable in that he delivers an amount that is smaller than the quantity ordered by the retailer. On the other hand, the backup supplier is perfectly reliable. At the beginning of the period the retailer places an order of Q units to the primary supplier and reserves K units with the backup supplier at a price of c_r per unit. We assume that the retailer pays only for the portion of the order delivered by the primary supplier, with purchase price c per unit. In addition, after the quantity delivered by the primary supplier is observed, he has the option to buy up to K units from the backup supplier at a price of c_e per unit. This option may be exercised before or after the realization of the demand, giving rise to two different models (Model 1 and 2, respectively). After the orders are delivered by both suppliers the retailer sells as much as possible at a price r per unit. Unsold products are salvaged at a price h per unit and unmet demand incurs a penalty p per unit. Our objective is to determine the order and reservation quantities that maximize the expected profit of the retailer.

We assume that demand X is a continuous random variable with probability density function f and differentiable cumulative distribution function F. Furthermore, we make the following assumptions regarding the cost and revenue parameters of the problem. First, the backup supplier is more expensive than the primary, that is, $c_r + c_e > c$, because otherwise the problem would reduce to the classical newsvendor problem with purchase price equal to $c_r + c_e$. Next, the benefit from the sale of one unit (revenue plus penalty saved) exceeds the purchase cost from either supplier, that is, $r + p > c_r + c_e$, so that a profit is possible. Finally, excess inventory is sold at a lower price than its purchase price from either supplier, that is, $h < \min\{c, c_e\}$. Violating this last assumption would lead to models that are either much simpler to analyze or already studied. Specifically, if $h > c$, the retailer would order as much as possible from the primary supplier and we would have a problem with one decision variable, the reservation quantity. Similarly, if $h > c_e$, the retailer would buy the whole quantity that has been reserved, making the reservation irrelevant because he would actually have to decide how much to buy from the reliable supplier at a price of $c_r + c_e$ per unit. Thus, we would have a special case of the model with two suppliers that has been studied by Wang et al. [84].

In the following sections we derive expressions for the objective function for Models 1 and 2 as well as some preliminary results that will be used in subsequent sections.

3.3.1 Model 1

In this model the retailer exercises the option to buy from the backup supplier before demand is realized (see 3.1 for a graphical representation of the sequence of events for this model). Having already received the quantity delivered by the primary supplier, the problem of deciding how much to buy from the backup supplier is equivalent to a newsvendor problem with starting inventory and a maximum order quantity, which is the quantity that has been reserved with the backup supplier. It is easy to show that the optimal order quantity is such that the total inventory is as close as possible to a target value I which is the solution of the classical newsvendor problem with purchase price c_e, that is,

$$I = F^{-1}\left(\frac{r + p - c_e}{r + p - h}\right). \tag{3.3.1}$$

Therefore, the retailer makes no purchase if the delivered quantity is greater than I, exercises his option in full if the delivered and reservation quantities are such that I cannot be reached, and buys the quantity needed to reach I otherwise. Because the maximum quantity that may be bought from the backup supplier is I, it is clear that $K > I$ cannot be optimal. Then, with S being the delivered quantity by the primary supplier, the expected profit is given by

$$\begin{aligned}\Pi(Q,K) &= -c_r K - cE(S) - c_e E\left[\min\{(I-S)^+, K\}\right] \\ &\quad + E\left[L\left(S + \min\{(I-S)^+, K\}\right)\right], \quad K \le I,\end{aligned} \tag{3.3.2}$$

where

$$L(z) = \int_0^z [rx + h(z-x)]f(x)dx + \int_z^\infty [rz - p(x-z)]f(x)dx$$

is the expected net revenue when the total quantity delivered by the two suppliers is equal to z.

Note that when only the backup supplier is available ($Q = 0$), we get from (3.3.2) that $\Pi(0, K) = -(c_r + c_e)K + L(K)$, which is the expected profit for the classical newsvendor problem with purchase price $c_r + c_e$. This is reasonable because it would not be optimal to reserve a quantity and then not to buy the whole of it. Then, denoting this quantity by K^*, we have

$$K^* = F^{-1}\left(\frac{r + p - c_r - c_e}{r + p - h}\right), \tag{3.3.3}$$

which is the well known critical ratio for the classical newsvendor problem.

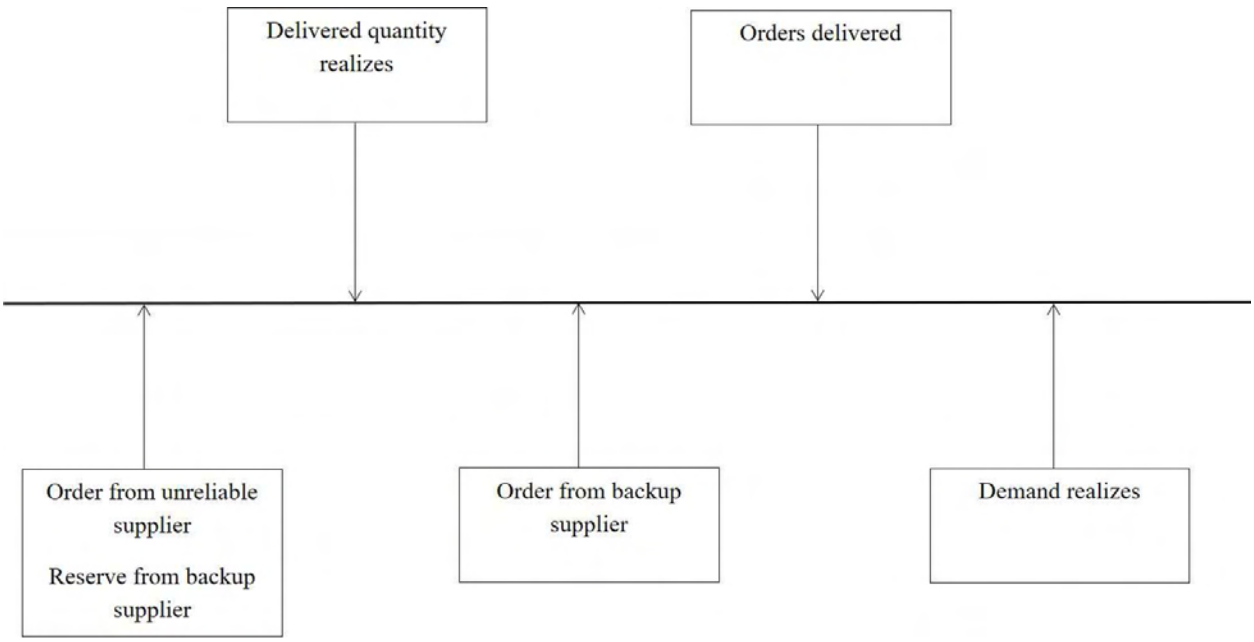

Figure 3.1: Sequence of events for Model 1

3.3.2 Model 2

In this model the retailer exercises the option to buy from the backup supplier after the demand becomes known, so he buys what is needed to satisfy the demand, subject to the restriction imposed by the quantity having been reserved. The sequence of events for this model is presented in Figure 3.2. If the demand is less than the delivered quantity there is no need to buy from the backup supplier and there is a surplus. If the demand is larger than the delivered quantity and can be covered by the backup supplier the retailer buys the quantity needed and there is neither surplus nor shortage. Finally, if the demand is larger than the delivered quantity but there is not enough reserved capacity to cover it, the retailer buys the whole reserved quantity and there is a shortage. Therefore, we get the following expression for the expected profit.

$$\Pi(Q,K) = -cE(S) - c_r K + E\left[L(S,K)\right], \tag{3.3.4}$$

where

$$\begin{aligned} L(z,k) &= \int\limits_{x=0}^{z} \left[rx + h(z-x)\right] f(x)dx + \int\limits_{x=z}^{z+k} \left[-c_e(x-z) + rx\right] f(x)dx \\ &+ \int\limits_{x=z+k}^{\infty} \left[-c_e k + r(z+k) - p(x-z-k)\right] f(x)dx, \end{aligned} \tag{3.3.5}$$

is the expected net revenue if the delivered quantity from the primary supplier is z and the reserved capacity with the backup supplier is k. Differentiating with respect to each argument we get

$$\frac{\partial L}{\partial z} = p + r - (c_e - h)F(z) - (p + r - c_e)F(z+k), \tag{3.3.6}$$

$$\frac{\partial L}{\partial k} = (p + r - c_e)\left[1 - F(z+k)\right], \tag{3.3.7}$$

which will be used later to get the first-order derivatives of the profit function.

3.4 Random yield

In this model the quantity delivered by the primary supplier is a random portion of the order placed by the retailer, that is, $S = QU$, where U is a random variable with support on [0,1]. For both Model 1 and 2 we determine properties of the optimal order and reservation quantities. For Model 2 in particular we also analyze models with two primary suppliers and two products.

3.4.1 Analysis of Model 1

We assume that random variable U is continuous with probability density function g and differentiable cumulative distribution function G. Setting $S = QU$ in (3.3.2) we obtain the following detailed expressions for the profit function. For $Q + K \leq I$,

$$\Pi(Q,K) = -(c_r + c_e)K - cQE(U) + E[L(QU + K)], \tag{3.4.1}$$

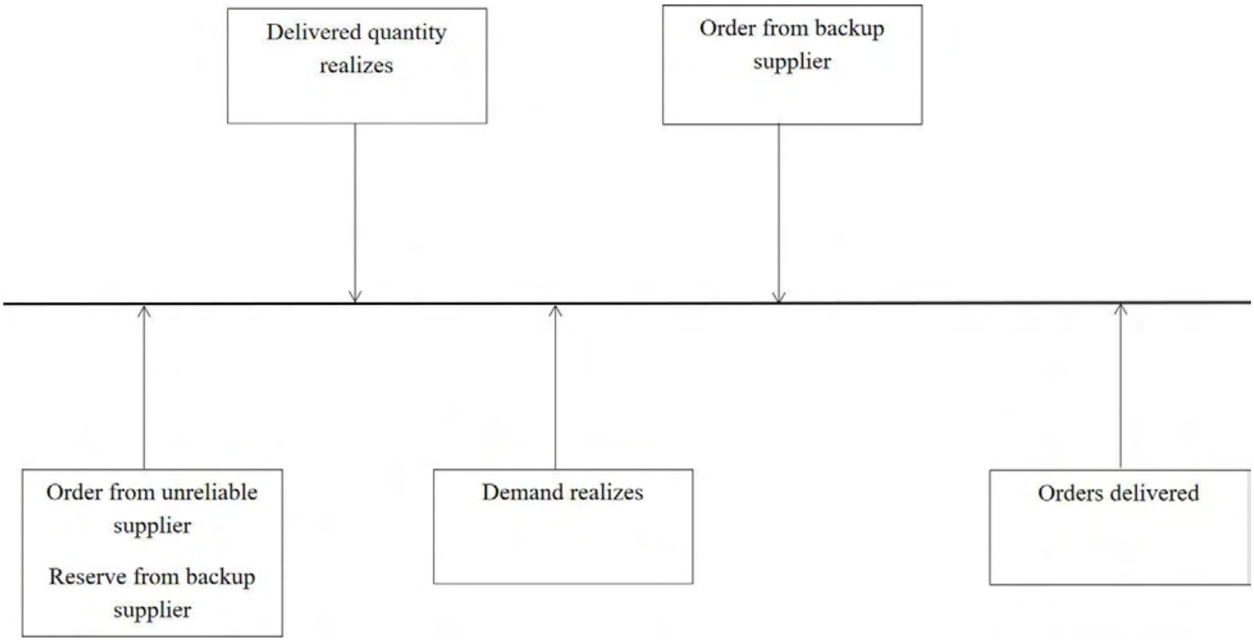

Figure 3.2: Sequence of events for Model 2

for $I - K < Q \leq I$,

$$\Pi(Q,K) = -c_r K - cQE(U) - c_e K G\left(\frac{I-K}{Q}\right) - c_e \int_{\frac{I-K}{Q}}^{1} (I - Qu) g(u) du + \int_{0}^{\frac{I-K}{Q}} L(Qu + K) g(u) du + L(I)\left[1 - G\left(\frac{I-K}{Q}\right)\right], \tag{3.4.2}$$

and for $Q > I$,

$$\Pi(Q,K) = -c_r K - cQE(U) - c_e K G\left(\frac{I-K}{Q}\right) - c_e \int_{\frac{I-K}{Q}}^{\frac{I}{Q}} (I - Qu) g(u) du + \int_{0}^{\frac{I-K}{Q}} L(Qu + K) g(u) du + L(I)\left[G\left(\frac{I}{Q}\right) - G\left(\frac{I-K}{Q}\right)\right] + \int_{\frac{I}{Q}}^{1} L(Qu) g(u) du. \tag{3.4.3}$$

Then, using the fact that $dL/dz = r + p - (r + p - h)F(z)$, we get the following for the first-order derivatives of the profit function.

For $Q + K \leq I$,

$$\frac{\partial \Pi}{\partial Q} = (r+p-c)E(U) - (r+p-h)E\left[UF(QU+K)\right], \tag{3.4.4}$$
$$\frac{\partial \Pi}{\partial K} = r+p-c_e-c_r-(r+p-h)E\left[F(QU+K)\right], \tag{3.4.5}$$

for $I - K < Q \leq I$,

$$\frac{\partial \Pi}{\partial Q} = -cE(U) + c_e \int_{\frac{I-K}{Q}}^{1} ug(u)du + \int_{0}^{\frac{I-K}{Q}} \left[r+p-(r+p-h)F(Qu+K)\right] ug(u)du, \tag{3.4.6}$$
$$\frac{\partial \Pi}{\partial K} = -c_r + (r+p-c_e)G\left(\frac{I-K}{Q}\right) - (r+p-h)\int_{0}^{\frac{I-K}{Q}} F(Qu+K)g(u)du, \tag{3.4.7}$$

and for $Q > I$,

$$\frac{\partial \Pi}{\partial Q} = -cE(U) + \int_{0}^{\frac{I-K}{Q}} \left[r+p-(r+p-h)F(Qu+K)\right] ug(u)du + c_e \int_{\frac{I-K}{Q}}^{\frac{I}{Q}} ug(u)du + \int_{\frac{I}{Q}}^{1} \left[r+p-(r+p-h)F(Qu)\right] ug(u)du, \tag{3.4.8}$$
$$\frac{\partial \Pi}{\partial K} = -c_r + (r+p-c_e)G\left(\frac{I-K}{Q}\right) - (r+p-h)\int_{0}^{\frac{I-K}{Q}} F(Qu+K)g(u)du. \tag{3.4.9}$$

In the following proposition, whose proof is given in Appendix C, we establish the concavity of the expected profit function.

Proposition 3.1. $\Pi(Q, K)$ *is jointly concave in* Q, K.

Because of Proposition 3.1 the KKT conditions, $\partial\Pi/\partial Q \leq 0$, $Q\left(\partial\Pi/\partial Q\right) = 0$, $\partial\Pi/\partial K \leq 0$, and $K\left(\partial\Pi/\partial K\right) = 0$, are necessary and sufficient for order and reservation quantities to be profit maximizing. We use these conditions to obtain expressions and bounds for these optimal quantities.

When only the primary supplier is available ($K = 0$), we get from (3.3.2) $\Pi(Q,0) = cQE(U) + E\left(L(QU)\right)$. Then, Proposition 3.1 implies that the optimal order quantity, denoted by Q^*, is obtained by setting the derivative of $\Pi(Q,0)$ equal to 0, leading to

$$E\left(UF(Q^*U)\right) = \frac{p+r-c}{p+r-h}E(U). \tag{3.4.10}$$

When only the backup supplier is available ($Q = 0$), the optimal reservation quantity K^* is given by (3.3.3). Note also that $K^* < Q^*$ because $K^* < Q_c$ and $Q_c \leq Q^*$, where $Q_c = F^{-1}((p+r-c)/(p+r-h))$ is the optimal order quantity to the primary supplier assuming he is perfectly reliable and the only one available.

It is reasonable that Q^* and K^* are upper bounds for the optimal order and reservation quantities when both suppliers are available. This is established in the following theorem along with a lower bound on the total order and reservation quantity.

Theorem 3.1. *Let Q_{o1}, K_{o1} be the optimal order and reservation quantities. Then, $Q_{o1} \leq Q^*$, $K_{o1} \leq K^*$, and $Q_{o1} + K_{o1} \geq K^*$.*

Proof. Because the profit function is concave we have $\partial\Pi/\partial Q < 0$ for $K = 0$ and $Q > Q^*$. Taking also into account that $\partial^2\Pi/\partial K\partial Q < 0$ (see proof of Proposition 3.1), we get that $\partial\Pi/\partial Q < 0$ for $Q > Q^*$ and any K. Therefore, $Q_{o1} \leq Q^*$ because otherwise the KKT conditions would not be satisfied. The proof of $K_{o1} \leq K^*$ uses similar arguments since $\partial\Pi/\partial K < 0$ for $Q = 0$ and $K > K^*$. To prove that $Q_{o1} + K_{o1} \geq K^*$ we only need to consider cases with $Q_{o1} < K^*$. With this in mind we define function

$$\tilde{\Pi}(q) = \left.\frac{\partial\Pi}{\partial K}\right|_{Q=q,K=K^*-q}, \quad 0 \leq q \leq K^*.$$

Because $K^* < I$, we use (3.4.5) to get

$$\frac{d\tilde{\Pi}}{dq} = (r+p-h)E\left[(1-U)f(qU + K^* - q)\right] \geq 0,$$

which combined with $\tilde{\Pi}(0) = 0$ yields

$$\left.\frac{\partial\Pi}{\partial K}\right|_{Q=Q_{o1},K=K^*-Q_{o1}} \geq 0.$$

Therefore, $K_{o1} \geq K^* - Q_{o1}$ because of the concavity of the profit function. □

The optimal policy is characterized in the next theorem.

Theorem 3.2. *Let $\tilde{c}_r = (r+p-h)\int_0^{\min\{1,I/Q^*\}}[F(I) - F(Q^*u)]g(u)du$. Then,*
i) If $c_r \geq \tilde{c}_r$, it is optimal not to reserve any capacity from the backup supplier and order Q^* from the primary supplier.
ii) If $c_r < \tilde{c}_r$, Q_{o1} and K_{o1} are positive and satisfy $\frac{\partial\Pi}{\partial Q} = 0, \frac{\partial\Pi}{\partial K} = 0$.

Proof. Part (i) is a consequence of the fact that the KKT conditions are satisfied by $K = 0$, $Q = Q^*$. Because the KKT conditions have to be satisfied by the optimal order and reservation quantities, for part (ii) it suffices to show that Q_{o1} and K_{o1} are positive. The proof is by contradiction. Because of Theorem 3.1 we only need to consider $Q \leq Q^*$ and $K \leq K^*$. For $Q = 0$ and $K \leq K^*$ we have from (3.4.4) and (3.3.3)

$$\frac{\partial \Pi}{\partial Q} \geq (c_e + c_r - c)E(U) > 0.$$

For $K = 0$ and $Q \leq Q^*$ we have from either (3.4.5) or (3.4.7), and (3.4.10)

$$\frac{\partial \Pi}{\partial K} \geq \tilde{c}_r - c_r > 0.$$

The KKT conditions are not satisfied, therefore $Q = 0$ and/or $K = 0$ cannot be optimal. □

In the following theorems we identify cases for which the area containing the optimal solution can be further restricted.

Theorem 3.3. *When $Q^* < I$, the optimal solution satisfies $Q_{o1} + K_{o1} \leq Q^*$.*

Proof. We define function

$$\bar{\Pi}(k) = \left.\frac{\partial \Pi}{\partial Q}\right|_{K=k, Q=Q^*-k}, \quad 0 \leq k \leq Q^*.$$

Because $Q + K = Q^* < I$, we use (3.4.4) to get

$$\frac{d\bar{\Pi}}{dk} = -(r + p - h)E\left[U(1-U)f(Q^*U + k(1-U))\right] \leq 0.$$

Taking also into account that $\bar{\Pi}(0) = 0$ and $K_{o1} \leq K^* < Q^*$ we get

$$\left.\frac{\partial \Pi}{\partial Q}\right|_{K=K_{o1}, Q=Q^*-K_{o1}} \leq 0,$$

and thus $Q_{o1} \leq Q^* - K_{o1}$ because of the concavity of the profit function. □

Based on Theorems 3.1 and 3.3 we provide in Figure 3.3 a graphical representation of the area containing the optimal solution when $Q^* < I$. It is interesting to note that in this case the retailer should buy the whole quantity that has been reserved with the backup supplier. Therefore, the problem is equivalent to one of optimal ordering from two suppliers, one subject to random yield and one perfectly reliable.

Next, we show that when the backup supplier remains more expensive than the primary supplier even after excluding reservation costs, it is optimal for the retailer to order at least I from the primary supplier because otherwise he would certainly have to buy some amount from the backup supplier at a higher price.

Theorem 3.4. *When $c_e \geq c$, the optimal solution satisfies $Q_{o1} \geq I$.*

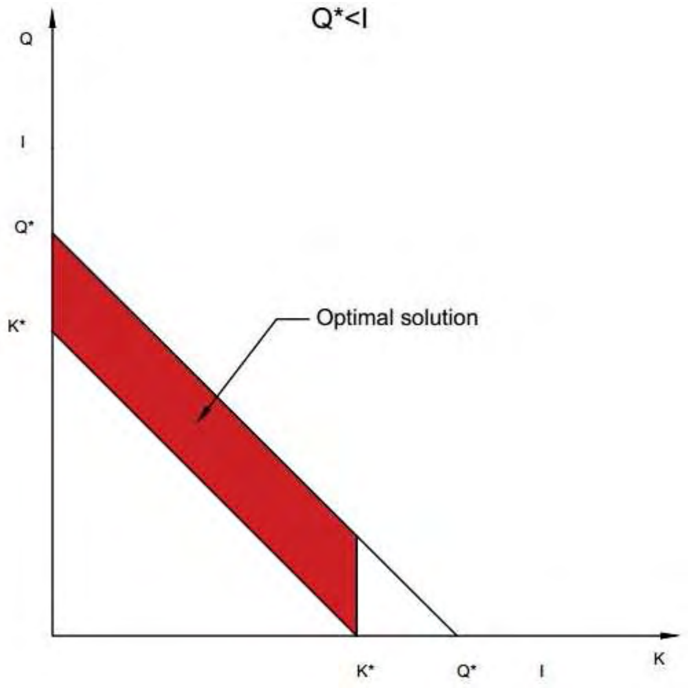

Figure 3.3: Area of the optimal solution for $Q^* < I$

Proof. Using (3.3.1), (3.4.6) can be written as

$$\frac{\partial \Pi}{\partial Q} = (c_e - c)E(U) + (r + p - h) \int_0^{\frac{I-K}{Q}} [F(I) - F(Qu + K)]\, ug(u)du \geq 0.$$

Therefore, $(\partial\Pi/\partial Q)|_{Q=I} \geq 0$, which implies that $Q_{o1} \geq I$ because of the concavity of the profit function. □

Theorems 3.1 and 3.4 imply that the optimal solution when $c_e \geq c$ is located within the circumscribed area of Figure 3.4. Note that Theorem 3.3 may not hold when $c_e \geq c$, in which case $Q^* > I$. This can happen because the reservation cost may be arbitrarily small resulting in the optimal reservation quantity being large enough so that $Q_{o1} + K_{o1} > Q^*$. This is illustrated in the following example.

Example 3.1. *Let X be normally distributed with mean 100 and standard deviation 30, U be uniformly distributed on [0,1], $r = 25$, $p = 15$, $h = 3$, $c_e = 9.7$, $c_r = 0.2$, and $c = 9.5$. Then, $Q^* = 249.09$, $Q_{o1} = 137.45$, and $K_{o1} = 115.04$, so the sum of the optimal order and reservation quantities is larger than Q^*.*

Finally, for cases with $c_e < c$ and $Q^* > I$ the optimal order and reservation quantities are constrained by Theorem 3.1 only (Figure 3.5), as we have not been able to obtain any additional results limiting the area containing the optimal solution. However, a numerical investigation that we conducted indicated that Theorem 3.3 may also hold in this case.

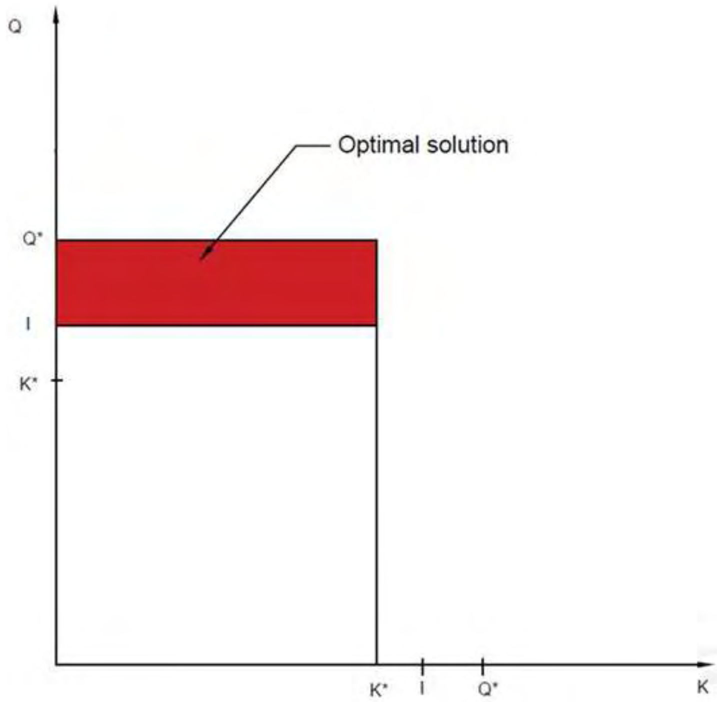

Figure 3.4: Area of the optimal solution for $c \leq c_e$

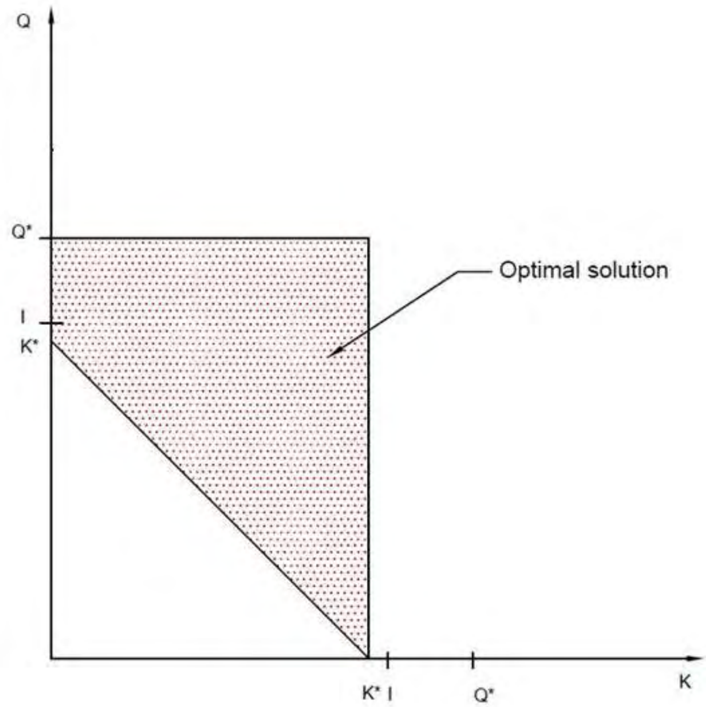

Figure 3.5: Area of the optimal solution for $c > c_e$ and $Q^* > I$

3.4.2 Analysis of Model 2

Unlike Model 1, the results of this model hold without the assumption that yield U is a continuous random variable. The expected profit when the retailer orders Q from the primary supplier and reserves K from the backup supplier is given by (3.3.4) with $S = QU$ and takes the form

$$\Pi(Q, K) = -cQE(U) - c_r K + E\left[L(QU, K)\right]. \tag{3.4.11}$$

The following proposition, which establishes the concavity of the profit function, has been proved by Giri and Bardhan [30] who studied the same model in the context of a coordination mechanism between the supplier and the retailer.

Proposition 3.2. $\Pi(Q, K)$ *is jointly concave in* Q, K*.*

Because of Proposition 3.2 the KKT conditions, $\partial\Pi/\partial Q \leq 0$, $Q\left(\partial\Pi/\partial Q\right) = 0$, $\partial\Pi/\partial K \leq 0$, and $K\left(\partial\Pi/\partial K\right) = 0$, are necessary and sufficient for order and reservation quantities to be profit maximizing. To derive expressions for the expected profit's first derivatives, we use (3.3.6)-(3.3.7) which combined with (3.4.11) yields

$$\begin{aligned} \frac{\partial \Pi}{\partial Q} &= (p + r - c)E(U) - (c_e - h)E\left[UF(QU)\right] \\ &\quad -(p + r - c_e)E\left[UF(QU + K)\right], && (3.4.12) \\ \frac{\partial \Pi}{\partial K} &= -c_r + (p + r - c_e)\left[1 - E\left(F(QU + K)\right)\right]. && (3.4.13) \end{aligned}$$

Let Q^*, K^* be the optimal order and reservation quantities when only the primary or the backup supplier is available; Q^* satisfies (3.4.10) and K^* is the value of K for which $\partial\Pi/\partial K = 0$ for $Q = 0$. Therefore, we get from (3.4.13)

$$K^* = F^{-1}\left(\frac{p + r - c_e - c_r}{p + r - c_e}\right). \tag{3.4.14}$$

The following theorem, which is actually a repetition of Theorem 3.1, provides bounds for the optimal order and reservation quantities when both suppliers are available.

Theorem 3.5. *Let* Q_{o2}, K_{o2} *be the optimal order and reservation quantities. Then,* $Q_{o2} \leq Q^*$*,* $K_{o2} \leq K^*$*, and* $Q_{o2} + K_{o2} \geq K^*$*.*

Proof. For any $Q > Q^*$ we have from (3.4.12) and (3.4.10)

$$\frac{\partial \Pi}{\partial Q} < (p + r - c)E(U) - (p + r - h)E\left[UF\left(Q^* U\right)\right] = 0.$$

Similarly, for any $K > K^*$ we have from (3.4.13) and (3.4.14)

$$\frac{\partial \Pi}{\partial K} < -c_r + (p + r - c_e)\left[1 - F\left(K^*\right)\right] = 0.$$

In both cases the KKT conditions are not satisfied, so $Q_{o2} \leq Q^*$ and $K_{o2} \leq K^*$. To prove that $Q_{o2}+K_{o2} \geq K^*$ we argue as in the proof of Theorem 3.1. Specifically, we define function $\tilde{\Pi}(q)$ in the same way and use (3.4.13) to get

$$\frac{d\tilde{\Pi}}{dq} = (p + r - c_e)E\left[(1 - U)f(qU + K^* - q)\right] \geq 0,$$

leading to $K_{o2} \geq K^* - Q_{o2}$. □

The next theorem characterizes the optimal policy.

Theorem 3.6. *Let* $\bar{c}_r = (p + r - c_e)\left[1 - E\left(F(Q^*U)\right)\right]$. *Then,*
i) If $c_r \geq \bar{c}_r$, *it is optimal not to reserve any capacity from the backup supplier and order* Q^* *from the primary supplier.*
ii) If $c_r < \bar{c}_r$, Q_{o2} *and* K_{o2} *are positive and satisfy*

$$\begin{aligned}
&(p + r - c_e)E\left[UF(Q_{o2}U + K_{o2})\right] + (c_e - h)E\left[UF(Q_{o2}U)\right] = (p + r - c)E(U),\\
&(p + r - c_e)E\left[F(Q_{o2}U + K_{o2})\right] = p + r - c_e - c_r.
\end{aligned}$$

Proof. Part (i) follows from the fact that the KKT conditions are satisfied by $K = 0, Q = Q^*$. As in the proof of Theorem 3.2, for part (ii) it suffices to show that Q_{o2} and K_{o2} are positive. Because of Theorem 3.5 we only need to consider $Q \leq Q^*$ and $K \leq K^*$. For $Q = 0$ and $K \leq K^*$ we have from (3.4.12) and (3.4.14)

$$\frac{\partial \Pi}{\partial Q} \geq (c_e + c_r - c)E(U) > 0.$$

For $K = 0$ and $Q \leq Q^*$ we have from (3.4.13) and (3.4.10)

$$\frac{\partial \Pi}{\partial K} \geq \bar{c}_r - c_r > 0.$$

The KKT conditions are not satisfied, therefore $Q = 0$ and/or $K = 0$ cannot be optimal. □

An incomplete form of Theorem 3.6 without the dependence of the optimal quantities on the reservation cost has been provided by Giri and Bardhan [30]. Although it is not an entirely novel result, we use Theorem 3.6 in the next section where we study a model with two primary suppliers.

3.4.2.1 Two primary suppliers

In this section we consider the case of a retailer that has access to two primary suppliers. The purchasing cost from supplier i, $i = 1, 2$, is c_i, where $c_i < c_r + c_e$. The yields of the two suppliers, denoted by U_1, U_2, are assumed to be independent random variables. Then, the expected profit when the retailer orders Q_1, Q_2 from the primary suppliers and reserves K with the backup supplier is given by

$$\Pi(Q_1, Q_2, K) = -c_1Q_1E(U_1) - c_2Q_2E(U_2) - c_rK + E\left[L(Q_1U_1 + Q_2U_2, K)\right]. \tag{3.4.15}$$

As was the case with one primary supplier, we show in the following proposition that the profit function is concave (see Appendix C for the proof).

Proposition 3.3. *$\Pi(Q_1, Q_2, K)$ is jointly concave in Q_1, Q_2, K.*

Because of Proposition 3.3, the KKT conditions, $\partial\Pi/\partial Q_i \le 0$, $Q_i\,(\partial\Pi/\partial Q_i) = 0$, $i = 1, 2$, $\partial\Pi/\partial K \le 0$, and $K\,(\partial\Pi/\partial K) = 0$, are necessary and sufficient for order and reservation quantities to be profit maximizing. Using (3.3.6) and (3.3.7) we get

$$\begin{aligned} \frac{\partial\Pi}{\partial Q_i} &= (p + r - c_i)E(U_i) - (c_e - h)E\left[U_i F(Q_1U_1 + Q_2U_2)\right] \\ &\quad -(p + r - c_e)E\left[U_i F(Q_1U_1 + Q_2U_2 + K)\right], \quad i = 1, 2, \qquad (3.4.16) \\ \frac{\partial\Pi}{\partial K} &= -c_r + (p + r - c_e)\left[1 - E\left(F(Q_1U_1 + Q_2U_2 + K)\right)\right]. \qquad (3.4.17) \end{aligned}$$

Let Q_1^*, Q_2^*, K^* be the optimal order and reservation quantities when only primary supplier 1, primary supplier 2, or the backup supplier is available. Then Q_i^*, $i = 1, 2$, satisfies

$$E\left[U_i F(Q_i^* U_i)\right] = \frac{(p + r - c_i)E(U_i)}{p + r - h}, \qquad (3.4.18)$$

and K^* is given by (3.4.14). The following theorem generalizes Theorem 3.5 for two suppliers.

Theorem 3.7. *Let Q_{1o}, Q_{2o}, K_o be the optimal order and reservation quantities. Then, $Q_{1o} \le Q_1^*$, $Q_{2o} \le Q_2^*$, $K_o \le K^*$, and $Q_{1o} + Q_{2o} + K_o \ge K^*$.*

Proof. Similarly to the proof of Theorem 3.5, we use (3.4.16)-(3.4.18) and (3.4.14) to show that for any $Q_1 > Q_1^*$, $Q_2 > Q_2^*$, and $K > K^*$, we have $\partial\Pi/\partial Q_1 < 0$, $\partial\Pi/\partial Q_2 < 0$, and $\partial\Pi/\partial K < 0$, respectively, so that $Q_{1o} \le Q_1^*$, $Q_{2o} \le Q_2^*$, and $K_o \le K^*$. To prove that $Q_{1o} + Q_{2o} + K_o \ge K^*$ we consider cases with $Q_{1o} + Q_{2o} < K^*$. For $0 \le q \le K^*$ and $0 \le t \le 1$ we define function

$$\tilde{\Pi}(q, t) = \left.\frac{\partial\Pi}{\partial K}\right|_{Q_1 = qt, Q_2 = (K^* - q)t, K = (1-t)K^*}. \qquad (3.4.19)$$

We have $\tilde{\Pi}(q, 0) = 0$ and from (3.4.17)

$$\begin{aligned} \frac{\partial\tilde{\Pi}}{\partial t} &= -(p + r - c_e)E\left\{\left[qU_1 + (K^* - q)U_2 - K^*\right]\right. \\ &\quad \left.\cdot f\left[qtU_1 + (K^* - q)tU_2 + (1 - t)K^*\right]\right\} \ge 0, \end{aligned}$$

because $0 \le q \le K^*$ and $U_1, U_2 \le 1$. Therefore, $\tilde{\Pi}(q, t) \ge 0$ for every $0 \le q \le K^*$ and $0 \le t \le 1$, which for $t = (Q_{1o} + Q_{2o})/K^*$ and $q = Q_{1o}K^*/(Q_{1o} + Q_{2o})$ yields

$$\left.\frac{\partial\Pi}{\partial K}\right|_{Q_1 = Q_{1o}, Q_2 = Q_{2o}, K = K^* - Q_{1o} - Q_{2o}} \ge 0,$$

which implies that $K_o \ge K^* - Q_{1o} - Q_{2o}$ because of the concavity of the profit function. □

For the problem with no backup supplier it is known that it is optimal to place an order to at least the cheaper supplier. We show in the following lemma that ordering from at least the cheaper supplier remains optimal when there is a backup supplier as well.

Lemma 3.1. *Assuming that $c_1 \ge c_2$, $Q_2 = 0$ cannot be optimal.*

Proof. The proof is by contradiction. Assuming that it is optimal not to place an order to the cheaper supplier, Theorem 3.6 implies that the KKT conditions for Q_1, K are satisfied by Q_{1o}, K_o such that

$$\begin{aligned}(p+r-c_1)E(U_1) &= (p+r-c_e)E\left[U_1F(Q_{1o}U_1+K_o)\right]+(c_e-h)E\left[U_1F(Q_{1o}U_1)\right] \\ &> E(U_1)\left\{(p+r-c_e)E\left[F(Q_{1o}U_1+K_o)\right]+(c_e-h)E\left[F(Q_{1o}U_1)\right]\right\},\end{aligned}$$

where the inequality follows from the fact that U_1 and $F(Q_{1o}U_1)$ are positively correlated random variables. Therefore,

$$(p+r-c_e)E\left[F(Q_{1o}U_1+K_o)\right]+(c_e-h)E\left[F(Q_{1o}U_1)\right] < p+r-c_1. \tag{3.4.20}$$

Setting now $Q_1 = Q_{1o}$, $Q_2 = 0$, and $K = K_o$ in (3.4.16) and taking into account the fact that U_1, U_2 are independent and (3.4.20) we get

$$\begin{aligned}\frac{\partial \Pi}{\partial Q_2} &= E(U_2)\left\{p+r-c_2-(p+r-c_e)E\left[F(Q_{1o}U_1+K_o)\right]\right. \\ &\quad \left.-(c_e-h)E\left[F(Q_{1o}U_1)\right]\right\} > E(U_2)(c_1-c_2) \geq 0,\end{aligned}$$

so the KKT condition is not satisfied. □

Let Q_{2b}, K_{2b} be the optimal order and reservation quantities when only regular supplier 2 is available and $c_r < c_{r1} = (p+r-c_e)\left[1-E\left(F(Q_2^*U_2)\right)\right]$. Then, according to Theorem 3.6, Q_{2b} and K_{2b} are positive and satisfy

$$\begin{aligned}&(p+r-c_e)E\left[U_2F(Q_{2b}U_2+K_{2b})\right] \\ &\quad +(c_e-h)E\left[U_2F(Q_{2b}U_2)\right] = (p+r-c_2)E(U_2),\end{aligned} \tag{3.4.21}$$

$$(p+r-c_e)E\left[F(Q_{2b}U_2+K_{2b})\right] = p+r-c_e-c_r. \tag{3.4.22}$$

Next, let $\bar{Q}_1, \bar{Q}_2$ be the optimal order quantities when there is no backup supplier and $c_2 \leq c_1 < c_{11} = p+r-(p+r-h)E\left[F(Q_2^*U_2)\right]$. Then, we can show by a straightforward application of the KKT conditions that $\bar{Q}_1$ and $\bar{Q}_2$ are positive and satisfy

$$E\left[U_iF(\bar{Q}_1U_1+\bar{Q}_2U_2)\right] = \frac{(p+r-c_i)E(U_i)}{p+r-h}, \quad i=1,2. \tag{3.4.23}$$

Note that similar expressions have been derived by Parlar and Wang [62] for a model where the retailer pays the primary suppliers for the whole quantities ordered.

Finally, we define two more critical values for c_1 and c_r. For $c_r < c_{r1}$ we let $c_{12} = c_r + c_e - (c_e-h)E\left[F(Q_{2b}U_2)\right]$ and for $c_1 < c_{11}$ we let $c_{r2} = (p+r-c_e)\left[1-E(F(\bar{Q}_1U_1+\bar{Q}_2U_2))\right]$. Then, taking also into account Lemma 3.1, the optimal ordering and reservation quantities are given in the following theorem.

Theorem 3.8. *Let $c_1 \geq c_2$. Then,*
i) If $c_1 \geq c_{11}$ and $c_r \geq c_{r1}$, we have $Q_{1o} = K_o = 0$ and $Q_{2o} = Q_2^*$.
ii) If $c_r < c_{r1}$ and $c_1 \geq c_{12}$, we have $Q_{1o} = 0$, $Q_{2o} = Q_{2b}$, and $K_o = K_{2b}$.

iii) If $c_1 < c_{11}$ and $c_r \geq c_{r2}$, we have $K_o = 0$, $Q_{1o} = \bar{Q}_1$, and $Q_{2o} = \bar{Q}_2$.
iv) Otherwise, Q_{1o}, Q_{2o}, and K_o are positive and satisfy

$$\begin{aligned}
&(p + r - c_e)E\left[U_i F(Q_{1o}U_1 + Q_{2o}U_2 + K_o)\right] \\
&\quad +(c_e - h)E\left[U_i F(Q_{1o}U_1 + Q_{2o}U_2)\right] = (p + r - c_i)E(U_i), \quad i = 1, 2, \\
&(p + r - c_e)E\left[F(Q_{1o}U_1 + Q_{2o}U_2 + K_o)\right] = p + r - c_e - c_r.
\end{aligned}$$

Proof. For parts (i)-(iii) it is straightforward to verify the KKT conditions. For part (iv), which implies that the optimal order and reservation quantities satisfy $\partial\Pi/\partial Q_i = 0$, $i = 1, 2$, and $\partial\Pi/\partial K = 0$, it suffices to show that Q_{1o} and K_o are positive. We do that by contradiction. Assume first that $Q_{1o} = K_o = 0$, in which case $Q_{2o} = Q_2^*$. Then, we get $\partial\Pi/\partial Q_1 = E(U_1)(c_{11} - c_1)$ and $\partial\Pi/\partial K = c_{r1} - c_r$, which violate the KKT conditions because $c_1 < c_{11}$ and/or $c_r < c_{r1}$. Next, if we assume that $Q_{1o} = 0$ and $K_o > 0$, it is necessary that $c_r < c_{r1}$, $Q_{2o} = Q_{2b}$, and $K_o = K_{2b}$ for the KKT conditions to be satisfied. However, in that case we get $\partial\Pi/\partial Q_1 = E(U_1)(c_{12} - c_1) > 0$. We are also led to a contradiction if we assume that $Q_{1o} > 0$ and $K_o = 0$, in which case we get $\partial\Pi/\partial K = c_{r2} - c_r > 0$. □

We end this section with a discussion on the critical values for c_1 and c_r. If for certain values of the problem parameters it is optimal not to place an order to supplier 1, it is reasonable that this would be also the case if the reservation cost decreases. Therefore, we make the conjecture that c_{12} is increasing with c_r and $c_{12} < c_{11}$ for $c_r < c_{r1}$. Based on a similar reasoning, we also make the conjecture that c_{r2} is increasing with c_1 and $c_{r2} < c_{r1}$ for $c_1 < c_{11}$. In Figure 3.6 we present a partition of the c_1-c_r space resulting from the optimal policy, where each region corresponds to the use or not of supplier 1 and the backup supplier. This particular partition was obtained with $r = 30$, $h = 3$, $p = 20$, $c_2 = 10$, $c_e = 13$, $X \sim N(100, 30)$, and $U \sim U(0, 1)$.

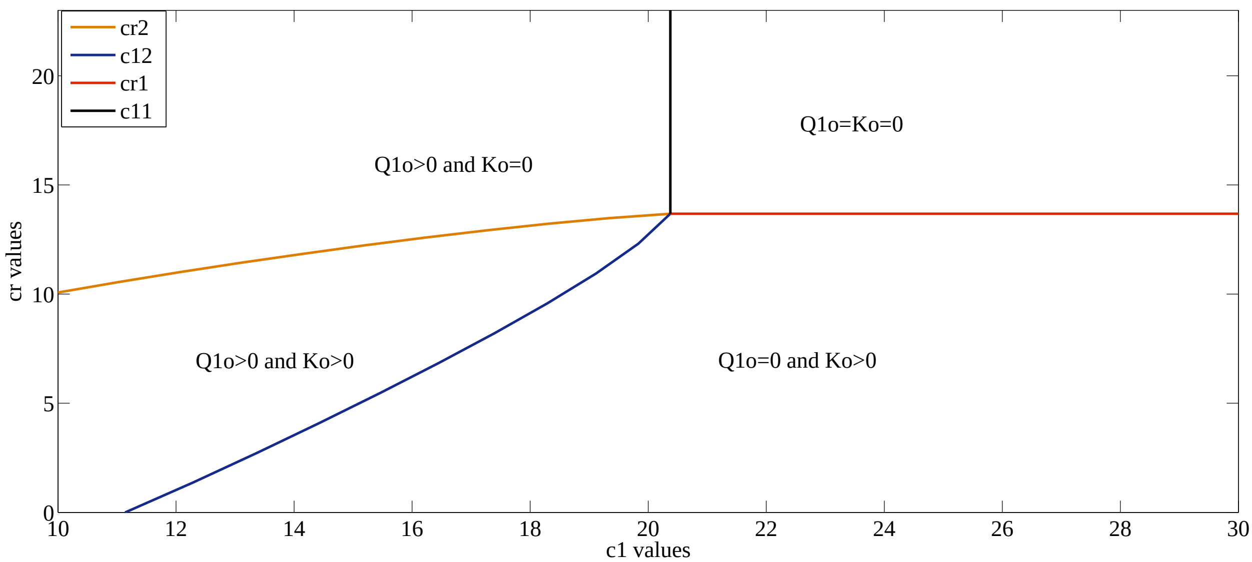

Figure 3.6: Optimal use of supplier 1 and backup supplier

3.4.2.2 Two products

In this section we consider a model with a retailer selling two products. There is one primary supplier for each product and a backup supplier that can supply both products. The retailer places orders to the primary suppliers and reserves a quantity with the backup supplier. After the demands of the two products become known the retailer has the right to buy any quantity of each product provided that the total does not exceed the quantity that has been reserved. We assume that the demands of the two products, denoted by X_1, X_2, are independent continuous random variables with probability density functions $f_i(x)$ and cumulative distribution functions $F_i(x)$, $i = 1, 2$, and the two suppliers have independent yields U_1, U_2. We let $c_i, r_i, p_i, h_i, c_{ei}$, $i = 1, 2$, be the purchase cost from the primary supplier, the retail price, the shortage cost, the salvage price, and the purchase cost from the backup supplier for product i. We denote by $L_i(z, k)$, $i = 1, 2$, the expected net revenue from product i if the delivered quantity from the primary supplier is z and the available reserved capacity of the backup supplier is k. Then, as in (3.3.5), we have

$$\begin{aligned} L_i(z,k) &= \int_{x=0}^{z} \left[r_i x + h_i(z-x)\right] f_i(x)dx + \int_{x=z}^{z+k} \left[-c_{ei}(x-z) + r_i x\right] f_i(x)dx \\ &+ \int_{x=z+k}^{\infty} \left[-c_{ei}k + r_i(z+k) - p_i(x-z-k)\right] f_i(x)dx. \end{aligned} \tag{3.4.24}$$

If the demands of both products exceed the delivered quantities from the primary suppliers, a decision has to be made on how the reserved capacity should be split between the two products. It is easy to see that priority should be given to the product for which the profit from a unit sale is larger, that is, first satisfy the demand of product $i = \arg\max_{j=1,2}(p_j + r_j - c_{ej})$. Without loss of generality we assume that $p_1 + r_1 - c_{e1} \geq p_2 + r_2 - c_{e2}$, so that product 1 has priority in the use of the backup supplier. Then, the expected profit when the retailer orders Q_1, Q_2 from the primary suppliers and reserves K with the backup supplier is given by

$$\Pi(Q_1, Q_2, K) = -c_1 Q_1 E(U_1) - c_2 Q_2 E(U_2) - c_r K + \Pi_1(Q_1, K) + \Pi_2(Q_1, Q_2, K), \tag{3.4.25}$$

where Π_i, $i = 1, 2$, is the expected profit from product i after the purchase and reservation costs have been paid. Note that Π_1 does not depend on Q_2 because the delivered quantity of product 2, which depends on Q_2, does not affect the amount of the reserved capacity used to satisfy the demand of product 1. Therefore,

$$\Pi_1(Q_1, K) = E\left[L_1(Q_1 U_1, K)\right]. \tag{3.4.26}$$

The available reserved capacity for product 2 is K if the demand of product 1 is fully satisfied by its primary supplier ($X_1 \leq Q_1 U_1$), $K - (X_1 - Q_1 U_1)$ if the demand of product 1 is partially satisfied and the reserved capacity is enough to cover it ($Q_1 U_1 < X_1 \leq Q_1 U_1 + K$), and 0 if the delivered quantity from the primary supplier together with the total reserved capacity are not enough to satisfy the demand of product 1 ($X_1 > Q_1U_1 + K$). Therefore,

$$\begin{aligned}\Pi_2(Q_1, Q_2, K) &= E\left[F_1(Q_1U_1)L_2(Q_2U_2, K)\right] \\ &+ E\left[\int\limits_{x=Q_1U_1}^{Q_1U_1+K} L_2(Q_2U_2, K - x + Q_1U_1)f_1(x)dx\right] \\ &+ E\left[(1 - F_1(Q_1U_1 + K))L_2(Q_2U_2, 0)\right]. \end{aligned} \tag{3.4.27}$$

The concavity of the profit function is established in the following proposition (for the proof see Appendix C).

Proposition 3.4. $\Pi(Q_1, Q_2, K)$ *is jointly concave in* Q_1, Q_2, K*.*

Because of Proposition 3.4, the KKT conditions, $\partial\Pi/\partial Q_i \le 0$, $Q_i\left(\partial\Pi/\partial Q_i\right) = 0$, $i = 1, 2$, $\partial\Pi/\partial K \le 0$, and $K\left(\partial\Pi/\partial K\right) = 0$, are necessary and sufficient for order and reservation quantities to be profit maximizing. Adjusting (3.3.6) and (3.3.7) to the case of two products, we get from (3.4.25)-(3.4.27)

$$\begin{aligned}\frac{\partial\Pi}{\partial Q_1} &= (p_1 + r_1 - c_1)E(U_1) - (c_{e1} - h_1)E\left[U_1F_1(Q_1U_1)\right] \\ &\quad -(p_1 + r_1 - c_{e1})E\left[U_1F_1(Q_1U_1 + K)\right] - (p_2 + r_2 - c_{e2}) \\ &\quad \times E\left[U_1F_1(Q_1U_1) - U_1F_1(Q_1U_1 + K) + U_1J(U_1, U_2)\right], \qquad (3.4.28) \\ \frac{\partial\Pi}{\partial Q_2} &= (p_2 + r_2 - c_2)E(U_2) - (p_2 + r_2 - h_2)E\left[U_2F_2(Q_2U_2)\right] \\ &\quad -(p_2 + r_2 - c_{e2})E\left[U_2F_1(Q_1U_1)F_2(Q_2U_2 + K)\right. \\ &\quad \left. -U_2F_1(Q_1U_1 + K)F_2(Q_2U_2) + U_2J(U_1, U_2)\right], \qquad (3.4.29) \\ \frac{\partial\Pi}{\partial K} &= -c_r + (p_1 + r_1 - c_{e1})\left\{1 - E\left[F_1(Q_1U_1 + K)\right]\right\} - (p_2 + r_2 - c_{e2}) \\ &\quad \times E\left[-F_1(Q_1U_1 + K) + F_1(Q_1U_1)F_2(Q_2U_2 + K) + J(U_1, U_2)\right], \qquad (3.4.30)\end{aligned}$$

where

$$J(U_1, U_2) = \int\limits_{Q_1U_1}^{Q_1U_1+K} F_2(Q_2U_2 + K - x + Q_1U_1)f_1(x)dx. \tag{3.4.31}$$

Let Q_1^*, Q_2^* be the optimal order quantities with no backup supplier and K^* be the optimal reservation quantity with no primary suppliers. Then Q_i^*, $i = 1, 2$, satisfy (3.4.18) and K^* can be obtained from $(\partial\Pi/\partial K)|_{Q_1=Q_2=0} = 0$, leading to

$$\begin{aligned}p_1 + r_1 - c_{e1} - c_r &= (p_1 + r_1 - c_{e1})F_1(K^*) \\ &\quad -(p_2 + r_2 - c_{e2})\left[F_1(K^*) - \int\limits_0^{K^*} F_2(K^* - x)f_1(x)dx\right]. \end{aligned} \tag{3.4.32}$$

The following theorem is a repetition of Theorem 3.7 for two suppliers and its proof follows similar steps.

Theorem 3.9. *Let Q_{1o}, Q_{2o}, K_o be the optimal order and reservation quantities. Then, $Q_{1o} \leq Q_1^*$, $Q_{2o} \leq Q_2^*$, $K_o \leq K^*$, and $Q_{1o} + Q_{2o} + K_o \geq K^*$.*

Proof. We prove that $Q_{1o} \leq Q_1^*$, $Q_{2o} \leq Q_2^*$, and $K_o \leq K^*$ by showing that for any $Q_1 > Q_1^*$, $Q_2 > Q_2^*$, and $K > K^*$ we have $\partial\Pi/\partial Q_1 < 0$, $\partial\Pi/\partial Q_2 < 0$, and $\partial\Pi/\partial K < 0$, respectively. From (3.4.28) and (3.4.29) we get

$$
\begin{aligned}
\frac{\partial^2\Pi}{\partial K \partial Q_1} &= -\left[(p_1 + r_1 - c_{e1}) - (p_2 + r_2 - c_{e2})\right] E\left[U_1 f_1(Q_1 U_1 + K)\right] \\
&\quad -(p_2 + r_2 - c_{e2}) E\left[U_1 f_1(Q_1 U_1 + K) F_2(Q_2 U_2) + U_1 \tilde{J}(U_1, U_2)\right], \\
\frac{\partial^2\Pi}{\partial K \partial Q_2} &= -(p_2 + r_2 - c_{e2}) E\left[U_2 F_1(Q_1 U_1) f_2(Q_2 U_2 + K) + U_2 \tilde{J}(U_1, U_2)\right],
\end{aligned}
$$

where

$$
\tilde{J}(U_1, U_2) = \int_{Q_1 U_1}^{Q_1 U_1 + K} f_2(Q_2 U_2 + K - x + Q_1 U_1) f_1(x) dx.
$$

Therefore, $\partial^2\Pi/(\partial K \partial Q_i) < 0$, $i = 1, 2$. We also have $\partial^2\Pi/\partial Q_i^2 < 0$ by the concavity of the profit function. Taking also into account that $\partial\Pi/\partial Q_i = 0$ for $Q_i = Q_i^*$, $K = 0$ and any value of Q_j, $j \neq i$, we conclude that $\partial\Pi/\partial Q_i < 0$ for any $Q_i > Q_i^*$. Similarly, we get $\partial\Pi/\partial K < 0$ for $K > K^*$ because $\partial\Pi/\partial K = 0$ for $K = K^*$ and $Q_1 = Q_2 = 0$, $\partial^2\Pi/\partial K^2 < 0$, and $\partial^2\Pi/(\partial Q_i \partial K) < 0$, $i = 1, 2$. To prove that $Q_{1o} + Q_{2o} + K_o \geq K^*$ it suffices to show that $\partial\tilde{\Pi}/\partial t \geq 0$, where function $\tilde{\Pi}(q, t)$ is defined in (3.4.19). With $\tilde{Q}_1 = qt$, $\tilde{Q}_2 = (K^* - q)t$, and $\tilde{K} = (1 - t)K^*$, we get from (3.4.30)

$$
\begin{aligned}
\frac{\partial\tilde{\Pi}}{\partial t} &= -\left[(p_1 + r_1 - c_{e1}) - (p_2 + r_2 - c_{e2})\right] (qU_1 - K^*) E\left[f_1(\tilde{Q}_1 U_1 + \tilde{K})\right] \\
&\quad -(p_2 + r_2 - c_{e2})(qU_1 - K^*) E\left[f_1(\tilde{Q}_1 U_1 + \tilde{K}) F_2(\tilde{Q}_2 U_2)\right] \\
&\quad -(p_2 + r_2 - c_{e2}) \left[(K^* - q)U_2 - K^*\right] E\left[F_1(\tilde{Q}_1 U_1) f_2(\tilde{Q}_2 U_2 + \tilde{K})\right] \\
&\quad -(p_2 + r_2 - c_{e2}) \left[(K^* - q)U_2 - K^* - qU_1\right] \\
&\quad \times E\left[\int_{\tilde{Q}_1 U_1}^{\tilde{Q}_1 U_1 + \tilde{K}} f_2(\tilde{Q}_2 U_2 + \tilde{K} - x + \tilde{Q}_1 U_1) f_1(x) dx\right] \geq 0,
\end{aligned}
$$

because $0 \leq q \leq K^*$ and $U_1, U_2 \leq 1$. □

We show in the following lemma that it is optimal to place a positive order to the primary supplier for product 1, which is reasonable because otherwise the demand for product 1 would have to be satisfied by the more expensive backup supplier.

Lemma 3.2. *$Q_1 = 0$ cannot be optimal.*

Proof. Setting $Q_1 = 0$ in (3.4.28) and (3.4.30) and taking into account that U_1, U_2 are independent we get

$$
\frac{1}{E(U_1)} \frac{\partial\Pi}{\partial Q_1} = p_1 + r_1 - c_1 + \frac{\partial\Pi}{\partial K} + c_r - (p_1 + r_1 - c_{e1}). \tag{3.4.33}
$$

Assume that $Q_1 = 0$ is optimal and let Q_{2o} and K_o be the optimal order quantity for product 2 and the optimal reservation quantity, respectively. Then, depending on whether $K_o > 0$ or $K_o = 0$, the KKT conditions imply that for $Q_2 = Q_{2o}$, $K = K_o$, $\partial\Pi/\partial K$ is either equal to 0 or $-c_r + p_1 + r_1 - c_{e1}$. Therefore, we get from (3.4.33)

$$\frac{1}{E(U_1)}\left.\frac{\partial\Pi}{\partial Q_1}\right|_{Q_1=0,Q_2=Q_{2o},K=K_o} = c_r + c_{e1} - c_1 \text{ or } p_1 + r_1 - c_1,$$

which is positive in both cases, leading to a contradiction. □

Regarding the optimal order to the primary supplier of product 2 and the optimal reservation quantity, it is clear that they cannot be both zero because in such a case there would be no supply source for product 2. Next, we define critical values for c_r and c_2 that determine whether the backup supplier should be used and whether an order should be placed to the primary supplier of product 2, respectively. We have

$$\begin{aligned}\hat{c}_r &= (p_1 + r_1 - c_{e1})\{1 - E[F_1(Q_1^*U_1)]\}\\ &\quad +(p_2 + r_2 - c_{e2})E[F_1(Q_1^*U_1)]\{1 - E[F_2(Q_2^*U_2)]\},\\ \hat{c}_2 &= p_2 + r_2 - (p_2 + r_2 - c_{e2})E[F_1(Q_{1b}U_1)]F_2(K_{1b})\\ &\quad -(p_2 + r_2 - c_{e2})E\left[\int_{Q_{1b}U_1}^{Q_{1b}U_1+K_{1b}} F_2(K_{1b} - x + Q_{1b}U_1)f_1(x)dx\right].\end{aligned}$$

where Q_{1b}, K_{1b} are the optimal order and reservation quantities when there is no regular supplier for product 2. It is clear that both are positive and they are obtained from (3.4.28) and (3.4.30) as the solution of $\partial\Pi/\partial Q_1 = 0$, $\partial\Pi/\partial K = 0$ with $Q_2 = 0$. Then, the optimal order and reservation quantities are given in the following theorem.

Theorem 3.10. *i) If $c_r \geq \hat{c}_r$, we have $K_o = 0$, $Q_{1o} = Q_1^*$, and $Q_{2o} = Q_2^*$.*
ii) If $c_2 \geq \hat{c}_2$, we have $Q_{1o} = Q_{1b}$, $Q_{2o} = 0$, and $K_o = K_{1b}$.
iii) Otherwise, Q_{1o}, Q_{2o} and K_o are positive and satisfy $\partial\Pi/\partial Q_i = 0$, $i = 1, 2$, and $\partial\Pi/\partial K = 0$.

Proof. For parts (i) and (ii) it is straightforward to verify the KKT conditions. For part (iii) it suffices to show that Q_{2o} and K_o are positive. Assuming that $Q_{2o} = 0$, it is necessary that $Q_{1o} = Q_{1b}$ and $K_o = K_{1b}$ for the KKT conditions to be satisfied. However, in that case we get $\partial\Pi/\partial Q_2 = E(U_2)(\hat{c}_2 - c_2) > 0$, violating the KKT conditions. We are also led to a contradiction if we assume that $K_o = 0$, in which case we get $\partial\Pi/\partial K = \hat{c}_r - c_r > 0$. □

According to part (ii) of the theorem, it may be optimal not to place an order to the primary supplier of product 2. This is surprising because the demand of product 2 would have to be exclusively satisfied by the more expensive backup supplier. To further elaborate on this matter we obtain two alternative expressions for $\hat{c}_2$. First, $\hat{c}_2$ can be written more compactly as

$$\hat{c}_2 = c_{e2} + (p_2 + r_2 - c_{e2})E\left[1 - \int_0^{Q_{1b}U_1+K_{1b}} F_2(K_{1b} + (Q_{1b}U_1 - x)^-)f_1(x)dx\right].$$

Second, using $(\partial \Pi / \partial K)|_{Q_1=Q_{1b}, Q_2=0, K=K_{1b}} = 0$, $\hat{c}_2$ can also be written as

$$\hat{c}_2 = c_r + c_{e2} + [(p_2 + r_2 - c_{e2}) - (p_1 + r_1 - c_{e1})] \{1 - E [F_1(Q_{1b}U_1 + K_{1b})]\},$$

so we get from the two expressions that $c_{e2} < \hat{c}_2 < c_r + c_{e2}$. The second inequality ensures that there exist feasible values of c_2 ($\hat{c}_2 \leq c_2 < c_r + c_{e2}$) for which no order is placed to the primary supplier of product 2, whereas the first inequality indicates that this may only happen if the primary supplier is more expensive than the backup supplier after excluding reservation costs. It seems that when c_{e2} is significantly smaller than c_2, it is preferable for the retailer to buy product 2 from the backup supplier after the demand for the high priority product 1 has been satisfied from the reserved quantity. Such a situation is illustrated in the following example.

Example 3.2. *Let X_1 be normally distributed with mean 100 and standard deviation 30, X_2 be normally distributed with mean 8 and standard deviation 2, U_1, U_2 be uniformly distributed on [0,1], $r_1 = 42$, $p_1 = 32$, $c_1 = 12$, $h_1 = 3$, $c_{e1} = 13$, $r_2 = 20$, $p_2 = 9$, $h_2 = 2$, $c_{e2} = 10.5$, and $c_r = 9$. Then, for $c_2 > 15.2$ the retailer should not place an order to the primary supplier of product 2.*

3.4.3 Comparison of Models 1 and 2

For a certain range of the reservation cost we can show that when the option to buy from the backup supplier can be exercised after the demand becomes known (Model 2), it is optimal to order less from the primary supplier and reserve more with the backup supplier compared to Model 1. To prove this we need to compare the critical reservation costs that determine whether the backup supplier should be used or not.

Lemma 3.3. *The critical reservation cost is larger for Model 2.*

Proof. The critical reservation costs can be written in the following forms.

$$\tilde{c}_r = (p + r - c_e) \left[\int_0^{\min\{1, I/Q^*\}} \left(1 - \frac{F(Q^*u)}{F(I)}\right) g(u) du \right], \quad (3.4.34)$$

$$\bar{c}_r = (p + r - c_e) \left[\int_0^1 (1 - F(Q^*u)) \, g(u) du \right]. \quad (3.4.35)$$

Then, $\tilde{c}_r < \bar{c}_r$ follows from (3.4.34) and (3.4.35) because $F(I) < 1$ implies that the integrand in (3.4.34) is smaller than the integrand in (3.4.35). □

Using Lemma 3.3 we get the following comparison of the optimal order and reservation quantities for the two models.

Theorem 3.11. *For $c_r \geq \tilde{c}_r$ we have $K_{o2} \geq K_{o1}$ and $Q_{o1} \geq Q_{o2}$.*

Proof. For $c_r \geq \bar{c}_r$ the theorem is satisfied with equalities because $K_{o1} = K_{o2} = 0$ and $Q_{o1} = Q_{o2} = Q^*$. For $\tilde{c}_r \leq c_r < \bar{c}_r$ we have $K_{o1} = 0$ and $Q_{o1} = Q^*$. For Model 2, we have $Q_{o2} \leq Q^*$ and $K_{o2} > 0$ from Theorems (3.5) and (3.6), respectively. □

3.5 Random capacity

In this model the primary supplier may not deliver the whole quantity ordered by the retailer because of insufficient capacity. Therefore, the delivered quantity is equal to $\min\{Y, Q\}$, where Y is a random variable denoting the supplier's capacity. We assume that Y is continuous with probability density function g and differentiable cumulative distribution function G. We also let $\bar{Y}$ be the maximum capacity of the supplier.

For the problem with only an unreliable supplier with capacity of infinite support ($\bar{Y} = \infty$) Ciarallo et al. [18] showed that the optimal order quantity, which we denote by Q_c, is not affected by the uncertain capacity of the supplier, that is, Q_c is the optimal order quantity for the classical newsvendor problem given by $F(Q_c) = (r + p - c)/(r + p - h)$. For the case with a reliable supplier the optimality of Q_c follows from the concavity of the expected profit function. When the supplier is unreliable the profit function is no longer concave but it does have a unique maximum as it is increasing for $Q \leq Q_c$ and decreasing for $Q > Q_c$. When the capacity has finite support ($\bar{Y} < \infty$), in which case we do not consider order quantities larger than the maximum capacity, the aforementioned monotonicity properties of the profit function imply that the maximum is achieved by $\bar{Q} = \min\{Q_c, \bar{Y}\}$.

In the following sections we show that under certain conditions $\bar{Q}$ is the optimal order quantity for the model with a backup supplier as well. For these cases the optimal reservation quantity is determined independently, using the fact that the profit function is concave in K. In any other case we show that the profit function, which is apparently not jointly concave in Q and K, has a unique maximum and we determine the optimal solution by using first-order conditions.

3.5.1 Analysis of Model 1

Considering order quantities that do not exceed the supplier's maximum capacity ($Q \leq \bar{Y}$), we set $S = \min\{Y, Q\}$ in (3.3.2) to obtain the following detailed expressions for the profit function. For $Q \leq \min\{I - K, \bar{Y}\}$,

$$\begin{aligned} \Pi(Q,K) &= -(c_r + c_e)K + \int_0^Q [L(y+K) - cy] g(y) dy \\ &\quad + [L(Q+K) - cQ]\,[1 - G(Q)]\,, \end{aligned} \tag{3.5.1}$$

for $I - K \leq Q \leq \min\{I, \bar{Y}\}$,

$$\begin{aligned} \Pi(Q,K) &= -c_r K + \int_0^{I-K} [L(y+K) - cy - c_e K] g(y) dy \\ &\quad + \int_{I-K}^{Q} [L(I) - cy - c_e(I - y)] g(y) dy \\ &\quad + [L(I) - cQ - c_e(I - Q)]\,[1 - G(Q)]\,, \end{aligned} \tag{3.5.2}$$

and for $I \leq Q \leq \bar{Y}$,

$$\begin{aligned}\Pi(Q,K) &= -c_r K + \int_0^{I-K} [L(y+K) - cy - c_e K] g(y) dy \\ &+ \int_{I-K}^{I} [L(I) - cy - c_e(I-y)] g(y) dy \\ &+ \int_I^Q [L(y) - cy] g(y) dy + [L(Q) - cQ] [1 - G(Q)] . \end{aligned} \quad (3.5.3)$$

When only the back up supplier is available ($Q = 0$), the optimal reservation quantity is given by (3.3.3). Denoting this optimal reservation quantity by K_{c1} for this model and using the fact that $dL/dz = r + p - (r + p - h)F(z)$, we get the following for the first-order derivatives of the profit function.
For $Q \leq \min\{I - K, \bar{Y}\}$,

$$\frac{\partial \Pi}{\partial Q} = (r + p - h)[F(Q_c) - F(Q + K)][1 - G(Q)], \quad (3.5.4)$$

$$\frac{\partial \Pi}{\partial K} = (r + p - h) [F(K_{c1}) - E[F(S + K)]] , \quad (3.5.5)$$

for $I - K \leq Q \leq \min\{I, \bar{Y}\}$,

$$\frac{\partial \Pi}{\partial Q} = (c_e - c)[[1 - G(Q)], \quad (3.5.6)$$

for $I \leq Q \leq \bar{Y}$,

$$\frac{\partial \Pi}{\partial Q} = (r + p - h)[F(Q_c) - F(Q)][1 - G(Q)], \quad (3.5.7)$$

and for $I - K \leq Q \leq \bar{Y}$,

$$\frac{\partial \Pi}{\partial K} = -c_r + (r + p - h) \int_0^{I-K} [F(I) - F(y + K)] g(y) dy. \quad (3.5.8)$$

Having derived first-order conditions (3.5.5) and (3.5.8) we can show that we may further restrict the search for the optimal reservation quantity in the set $K < K_{c1}$, i.e., we only need to consider values that are smaller than the optimal quantity ordered when the reliable supplier is the only one available. To see this, let $K \geq K_{c1}$. Then, from (3.5.5) we get $\partial \Pi / \partial K < 0$, and from (3.5.8), (3.3.1) and (3.3.3)

$$\begin{aligned}\frac{\partial \Pi}{\partial K} &< -c_r + (r + p - h) \int_0^{I-K_{c1}} [F(I) - F(K_{c1})] g(y) dy \\ &= -c_r [1 - G(I - K_{c1})] \leq 0.\end{aligned}$$

In the remainder of the section we derive expressions for the optimal order and reservation quantities, denoted by Q_{o1} and K_{o1}, respectively. It turned out that these expressions depend on which supplier is more expensive (excluding reservation costs), so we present the two cases separately.
Case 1: $c_e \geq c$. In this case, for any quantity having been reserved, there is no incentive for the retailer to order less from the primary supplier in anticipation of buying from the backup supplier at a lower price. Therefore, the availability of the backup supplier does not affect the optimal order to the primary supplier. This is stated formally in the following theorem.

Theorem 3.12. $Q_{o1} = \bar{Q}$.

Proof. Because $F(Q+K) \leq F(I)$ in (3.5.4) and $c_e \geq c$, which implies $F(I) \leq F(Q_c)$, we get from (3.5.4) and (3.5.6) that $(\partial\Pi/\partial Q) \geq 0$ for $Q \leq \min\{I, \bar{Y}\}$ and any K. Therefore, when $\bar{Y} \leq I$, the optimal order quantity is $\bar{Y}$, which is equal to $\bar{Q}$ because $I \leq Q_c$. Taking also into account (3.5.7) we see that the same is true when $I < \bar{Y} \leq Q_c$ because $(\partial\Pi/\partial Q) > 0$ for any $Q < \bar{Y}$. Finally, when $Q_c < \bar{Y}$, we have $(\partial\Pi/\partial Q) \geq 0$ for $Q < Q_c$ (with equality only if $c_e = c$, in which case $Q_c = I$) and $(\partial\Pi/\partial Q) < 0$ for Q such that $Q_c < Q < \bar{Y}$. Therefore, Q_c is the optimal order quantity, which is equal to $\bar{Q}$. □

Theorem 3.12 implies that the optimal reservation quantity is given by $K_{o1} = \arg\max\{\tilde{\Pi}(K)\}$, where $\tilde{\Pi}(K) = \Pi(\bar{Q}, K)$; $\tilde{\Pi}(K)$ is concave, which is a consequence the following lemma.

Lemma 3.4. *$\Pi(Q, K)$ is concave in K.*

Proof. Differentiating (3.5.5) and (3.5.8) we get the following. For $Q \leq \min\{I - K, \bar{Y}\}$,

$$\frac{\partial^2\Pi}{\partial K^2} = -(r + p - h)\left[E[f(S + K)]\right],$$

and for $I - K \leq Q \leq \bar{Y}$,

$$\frac{\partial^2\Pi}{\partial K^2} = -(r + p - h)\int_0^{I-K} f(y + K)g(y)dy.$$

Therefore, $\partial^2\Pi/\partial K^2 \leq 0$ for any $Q \leq \bar{Y}$ and $K < K_{c1}$, which combined with the fact that $\partial\Pi/\partial K$ is continuous (the expressions in (3.5.5) and (3.5.8) are identical for $K = I - Q$), proves that the profit function is concave with respect to its second argument. □

Taking into account Lemma 3.4 we derive expressions for the optimal reservation quantity in the following two theorems. Note that different expressions are obtained for $\bar{Y} \geq I$ and $\bar{Y} < I$.

Theorem 3.13. *For $\bar{Y} \geq I$, let $c_{r1} = (r + p - h)\int_0^I [F(I) - F(y)]g(y)dy$. Then,*
i) If $c_r \geq c_{r1}$, it is optimal not to reserve any capacity from the backup supplier.
ii) If $c_r < c_{r1}$, K_{o1} satisfies $\int_0^{I-K_{o1}} [F(I) - F(y + K_{o1})]g(y)dy = c_r/(r + p - h)$.

Proof. In this case we have $\bar{Q} \geq I$ because $Q_c \geq I$ and $\bar{Y} \geq I$. Therefore, $d\tilde{\Pi}/dK$ is given by the righthand side of (3.5.8) and the result follows from $\tilde{\Pi}(K)$ being concave and $(d\tilde{\Pi}/dK)|_{K=0} \leq 0$ for $c_r \geq c_{r1}$. □

Theorem 3.14. *For $\bar{Y} < I$, let $c_{r2} = (r+p-h)[F(I) - E[F(Y+I-\bar{Y})]]$ and $c_{r3} = (r+p-h)[F(I) - E[F(Y)]]$. Then,*
i) If $c_r \geq c_{r3}$, it is optimal not to reserve any capacity from the backup supplier.
ii) If $c_{r2} \leq c_r < c_{r3}$, K_{o1} satisfies $E[F(Y + K_{o1})] = F(K_{c1})$.
iii) If $c_r < c_{r2}$, K_{o1} satisfies $\int_0^{I-K_{o1}}[F(I) - F(y+K_{o1})]g(y)dy = c_r/(r+p-h)$.

Proof. In this case we have $\bar{Y} < I \leq Q_c$, so that $\bar{Q} = \bar{Y}$. Then, $d\tilde{\Pi}/dK$ is equal to the righthand side of (3.5.5) (with $S = \min\{Y, \bar{Y}\} = Y$) and (3.5.8) for $K \leq I - \bar{Y}$ and $K \geq I - \bar{Y}$, respectively, and the result follows from $\tilde{\Pi}(K)$ being concave, $(d\tilde{\Pi}/dK)|_{K=0} \leq 0$ for $c_r \geq c_{r3}$, and $(d\tilde{\Pi}/dK)|_{K=I-\bar{Y}} \leq 0$ for $c_r \geq c_{r2}$. □

Case 2: $c_e < c$. When it is cheaper to buy from the backup supplier, the sum of the optimal order and reservation quantity is equal to the solution of the classical newsvendor problem, unless it is restricted by the maximum capacity of the primary supplier. In other words, a portion of the order that would have been placed with the primary supplier if he was the only one available, is reserved with the backup supplier. This is shown in the following Lemma.

Lemma 3.5. *Let $Q^*(K)$ be the optimal order quantity to the primary supplier assuming a quantity of K has been reserved with the backup supplier. Then, for $K < K_{c1}$, we have $Q^*(K) = \min\{Q_c - K, \bar{Y}\}$.*

Proof. Note first that $Q_c < I$ because $c_e < c$. Then, for fixed K, we get from (3.5.6) and (3.5.7) that $(\partial\Pi/\partial Q) \leq 0$ for $Q \geq I - K$, if applicable, that is, if $\bar{Y} \geq I - K$. Therefore, $Q^*(K) \leq \min\{I - K, \bar{Y}\}$, which means that it can be determined from (3.5.4). Because $K_{c1} < Q_c$, for $K < K_{c1}$ we have $(\partial\Pi/\partial Q)|_{Q=0} > 0$. Then, when $Q_c \leq \bar{Y}$, we have $Q^*(K) = Q_c - K$ because we see from (3.5.4) that the profit function is increasing for $Q \leq Q_c - K$ and decreasing for $Q_c - K \leq Q \leq \min\{I - K, \bar{Y}\}$. When $\bar{Y} < Q_c$, we use the same argument to get that $Q^*(K) = \bar{Y}$ for $K \leq Q_c - \bar{Y}$ and $Q^*(K) = Q_c - K$ otherwise. □

As a result of Lemma 3.5 we end up with an one-variable optimization problem because the optimal reservation quantity is given by $K_{o1} = \arg\max\{\bar{\Pi}(K)\}$, where $\bar{\Pi}(K) = \Pi(\min\{Q_c - K, \bar{Y}\}, K)$. Moreover, $\bar{\Pi}(K)$ is concave, a fact established in the following lemma.

Lemma 3.6. $\bar{\Pi}(K)$ *is concave.*

Proof. If $\bar{Y} < Q_c$ and $K \leq Q_c - \bar{Y}$ we get from (3.5.5)

$$\frac{d\bar{\Pi}}{dK} = \left.\frac{\partial\Pi}{\partial K}\right|_{Q=\bar{Y}} = (r+p-h)\left[F(K_{c1}) - E[F(Y+K)]\right], \tag{3.5.9}$$

because $S = \min\{Y, \bar{Y}\} = Y$. In any other case (3.5.5) yields

$$\frac{d\bar{\Pi}}{dK} = \frac{\partial \Pi}{\partial K}\bigg|_{Q=Q_c-K} = (r+p-h)\left[F(K_{c1}) - E[F(\min\{Y, Q_c - K\} + K)]\right]. \quad (3.5.10)$$

Differentiating (3.5.9) and (3.5.10) we get

$$\frac{d^2\bar{\Pi}}{dK^2} = -(r+p-h) \int_0^{\min\{Q_c-K,\bar{Y}\}} f(y+K)g(y)dy.$$

Therefore, $d^2\bar{\Pi}/dK^2 \le 0$ for any $K < K_{c1}$. The proof is completed by noting that $d\bar{\Pi}/dK$ is continuous (the expressions in (3.5.9) and (3.5.10) are identical for $K = Q_c - \bar{Y}$). ☐

In the following theorems we derive expressions for the optimal reservation quantity based on Lemma 3.6, and the optimal order to the primary supplier is then computed according to Lemma 3.5.

Theorem 3.15. *For $\bar{Y} \ge Q_c$, let $c_{r4} = c - c_e + (r+p-h)\int_0^{Q_c}[F(Q_c) - F(y)]g(y)dy$. Then,*
i) If $c_r \ge c_{r4}$, $K_{o1} = 0$, and $Q_{o1} = Q_c$.
ii) If $c_r < c_{r4}$, K_{o1} satisfies $\int_0^{Q_c-K_{o1}}[F(Q_c) - F(y+K_{o1})]g(y)dy = (c_r + c_e - c)/(r+p-h)$ and $Q_{o1} = Q_c - K_{o1}$.

Proof. When $\bar{Y} \ge Q_c$, $d\bar{\Pi}/dK$ is given by (3.5.10), which after some straightforward algebra takes the form

$$\frac{d\bar{\Pi}}{dK} = c - c_e - c_r + (r+p-h) \int_0^{Q_c-K} [F(Q_c) - F(y+K)]g(y)dy.$$

Then, the result follows from $\bar{\Pi}(K)$ being concave and $(d\bar{\Pi}/dK)|_{K=0} \le 0$ for $c_r \ge c_{r4}$. ☐

Theorem 3.16. *For $\bar{Y} < Q_c$, let $c_{r5} = (r+p-h)[F(I) - E[F(Y + Q_c - \bar{Y})]]$ and $c_{r6} = (r+p-h)[F(I) - E[F(Y)]]$. Then,*
i) If $c_r \ge c_{r6}$, $K_{o1} = 0$ and $Q_{o1} = \bar{Y}$.
ii) If $c_{r5} \le c_r < c_{r6}$, K_{o1} satisfies $E[F(Y + K_{o1})] = F(K_{c1})$ and $Q_{o1} = \bar{Y}$.
iii) If $c_r < c_{r5}$, K_{o1} satisfies $\int_0^{Q_c-K_{o1}}[F(Q_c) - F(y+K_{o1})]g(y)dy = (c_r + c_e - c)/(r+p-h)$ and $Q_{o1} = Q_c - K_{o1}$.

Proof. For $K \le Q_c - \bar{Y}$, $d\bar{\Pi}/dK$ is given by (3.5.9), which can be also written as

$$\frac{d\bar{\Pi}}{dK} = -c_r + (r+p-h)[F(I) - E[F(Y+K)]].$$

Then, the result follows from from $\bar{\Pi}(K)$ being concave, $(d\bar{\Pi}/dK)|_{K=0} \le 0$ for $c_r \ge c_{r6}$, and $(d\bar{\Pi}/dK)|_{K=Q_c-\bar{Y}} \le 0$ for $c_r \ge c_{r5}$. ☐

3.5.2 Analysis of Model 2

Setting $S = \min\{Y, Q\}$ in (3.3.4) we get for $Q \leq \bar{Y}$

$$\Pi(Q,K) = -c_r K + \int_0^Q [L(y,K) - cy]g(y)dy + \int_Q^{\bar{Y}} [L(Q,K) - cQ]g(y)dy. \qquad (3.5.11)$$

Using (3.5.11) and (3.3.6)-(3.3.7) we obtain the first-order derivatives of the profit function. We have

$$\frac{\partial \Pi}{\partial Q} = A(Q,K)[1 - G(Q)], \qquad (3.5.12)$$

where

$$A(Q,K) = (r + p - c) - (c_e - h)F(Q) - (r + p - c_e)F(Q + K), \qquad (3.5.13)$$

and

$$\frac{\partial \Pi}{\partial K} = r + p - c_r - c_e - (r + p - c_e)E\left[F(S + K)\right]. \qquad (3.5.14)$$

It is easy to see that $\partial^2\Pi/\partial K^2 < 0$, implying that the profit function is concave in its second argument. Then, the optimal reservation quantity when only the backup supplier is available is the solution of $(\partial\Pi/\partial K) = 0$ with $S = 0$. Denoting this quantity by K_{c2} we have

$$K_{c2} = F^{-1}\left(\frac{r + p - c_r - c_e}{r + p - c_e}\right), \qquad (3.5.15)$$

and (3.5.14) can be written as

$$\frac{\partial \Pi}{\partial K} = (r + p - c_e)\left[F(K_{c2}) - E\left[F(S + K)\right]\right], \qquad (3.5.16)$$

implying that $\partial\Pi/\partial K < 0$ for $K \geq K_{c2}$. Therefore, as was the case with Model 1, the optimal reservation quantity is bounded above by the optimal quantity reserved in the absence of a primary, unreliable supplier.

Because the profit function is not jointly concave in its arguments, we obtain the characterization of the optimal solution by studying the equivalent problem of maximizing profit function $\tilde{\Pi}(K) = \Pi(Q^*(K), K)$, $K < K_{c2}$, where $Q^*(K)$ is defined as in Lemma 3.5 and is determined from (3.5.12). Note that $A(Q,K)$ is decreasing in Q. Therefore, if $\partial\Pi/\partial Q < 0$ for some $Q = \tilde{Q}$, we have $\partial\Pi/\partial Q < 0$ for all Q such that $\tilde{Q} < Q < \bar{Y}$. This implies that for fixed K, $\Pi(Q,K)$ has a unique maximum. Furthermore, for $K < K_{c2}$ we get from (3.5.12), (3.5.13) and (3.5.15) that $(\partial\Pi/\partial Q)|_{Q=0} > c_r + c_e - c > 0$. Therefore, $Q^*(K) = \bar{Y}$ if $A(\bar{Y}, K) \geq 0$, and otherwise it is obtained by setting the derivative in (3.5.12) equal to 0, that is, $Q^*(K)$ satisfies

$$(c_e - h)F(Q^*(K)) + (r + p - c_e)F(Q^*(K) + K) = r + p - c. \qquad (3.5.17)$$

Then, taking also into account that $A(Q,K)$ is non-increasing in K, we get the following characterization of the optimal order given the reservation quantity.

Lemma 3.7. *i) If $A(\bar{Y},0) \leq 0$, $Q^*(K)$ satisfies* (3.5.17) *for all $K < K_{c2}$.*
ii) If $A(\bar{Y},K_{c2}) \geq 0$, $Q^*(K) = \bar{Y}$ for all $K < K_{c2}$.
iii) If $A(\bar{Y},0) > 0$ and $A(\bar{Y},K_{c2}) < 0$, $Q^*(K) = \bar{Y}$ for $K \leq K_1$ and $Q^*(K)$ satisfies (3.5.17) *for $K_1 < K < K_{c2}$, where K_1 is such that $A(\bar{Y},K_1) = 0$.*

In summary, Lemma 3.7 states that there exists K_2 such that $Q^*(K) = \bar{Y}$ for $K < K_2$ and $Q^*(K)$ satisfies (3.5.17) for $K \geq K_2$ ($K_2 = 0, K_{c2}, K_1$ in cases (i),(ii),(iii), respectively). We use this in the following lemma, where we prove the concavity of $\tilde{\Pi}(K)$.

Lemma 3.8. $\tilde{\Pi}(K)$ *is concave.*

Proof. We have

$$\frac{d\tilde{\Pi}}{dK} = \frac{\partial \Pi}{\partial K}\bigg|_{Q=Q^*(K)} + \frac{\partial \Pi}{\partial Q}\bigg|_{Q=Q^*(K)} \frac{dQ^*(K)}{dK} = \frac{\partial \Pi}{\partial K}\bigg|_{Q=Q^*(K)}, \tag{3.5.18}$$

because $dQ^*/dK = d\bar{Y}/dK = 0$ for $K < K_2$ and $(\partial\Pi/\partial Q)|_{Q=Q^*(K)} = 0$ for $K \geq K_2$. For $K < K_2$, in which case $Q^*(K) = \bar{Y}$, we have $d^2\tilde{\Pi}/dK^2 < 0$ because $\Pi(Q,K)$ is concave in K. For $K \geq K_2$ we set $Q = Q^*(K)$ in (3.5.14) and differentiate it to get

$$\begin{aligned} \frac{d^2\tilde{\Pi}}{dK^2} &= -(r+p-c_e)\left[\int_0^{Q^*} f(y+K)g(y)dy \right. \\ &\quad \left. + \left[1 + \frac{dQ^*}{dK}\right] f(Q^*+K)(1-G(Q^*))\right], \end{aligned} \tag{3.5.19}$$

where we have dropped the dependence of Q^* on K for notational convenience. Differentiating now both sides of (3.5.17) we get

$$1 + \frac{dQ^*}{dK} = \frac{(c_e-h)f(Q^*)}{(c_e-h)f(Q^*) + (r+p-c_e)f(Q^*+K)} > 0, \tag{3.5.20}$$

which combined with (3.5.19) yields $d^2\tilde{\Pi}/dK^2 \leq 0$ for $K \geq K_2$ as well. Because F is continuous, it is easy to see from (3.5.17) that $Q^*(K)$ is also continuous. Then, the continuity of $d\tilde{\Pi}/dK$ follows from (3.5.18) and (3.5.14), completing the proof. □

Having established by Lemma 3.8 that the profit function has a unique maximum, in the next two theorems we derive expressions for the optimal order and reservation quantities, denoted by Q_{o2} and K_{o2}, respectively.

Theorem 3.17. *For $\bar{Y} \geq Q_c$, let $c_{r7} = (r+p-c_e)\left[1 - E[F(\min\{Y,Q_c\})]\right]$. Then,*
i) If $c_r \geq c_{r7}$, $K_{o2} = 0$ and $Q_{o2} = Q_c$.
ii) If $c_r < c_{r7}$, Q_{o2} and K_{o2} satisfy $A(Q_{o2},K_{o2}) = 0$ and $E[F(\min\{Y,Q_{o2}\} + K_{o2})] = F(K_{c2})$.

Proof. When $\bar{Y} \geq Q_c$, we have from (3.5.13) that $A(\bar{Y},0) \leq 0$, which implies that $Q^*(K)$ is given by (3.5.17) (part (i) of Lemma 3.7). In particular, $Q^*(0) = Q_c$ so that

$$\frac{d\tilde{\Pi}}{dK}\bigg|_{K=0} = \frac{\partial \Pi}{\partial K}\bigg|_{Q=Q_c,K=0} = -c_r + c_{r7}.$$

Then, $(d\tilde{\Pi}/dK)|_{K=0} \leq 0$ for $c_r \geq c_{r7}$ and part (i) follows from Lemma 3.8. On the other hand, $(d\tilde{\Pi}/dK)|_{K=0} > 0$ for $c_r < c_{r7}$, so $(d\tilde{\Pi}/dK)|_{K=K_{o2}} = 0$. Then, noting that $Q^*(K_{o2}) = Q_{o2}$, part (ii) follows from (3.5.17), (3.5.18) and (3.5.16). □

Theorem 3.18. *For $\bar{Y} < Q_c$, let $c_{r8} = (r+p-c_e)\,[1 - E[F(Y)]]$ and $c_{r9} = (r+p-c_e)\,[1 - E[F(Y+K_1)]]\,\mathbf{1}[A(\bar{Y}, K_{c2}) < 0]$, where K_1 satisfies $A(\bar{Y}, K_1) = 0$ when $A(\bar{Y}, K_{c2}) < 0$. Then,*
i) If $c_r \geq c_{r8}$, $K_{o2} = 0$ and $Q_{o2} = \bar{Y}$.
ii) If $c_{r9} \leq c_r < c_{r8}$, K_{o2} satisfies $E[F(Y + K_{o2})] = F(K_{c2})$ and $Q_{o2} = \bar{Y}$.
iii) If $c_r < c_{r9}$, Q_{o2} and K_{o2} satisfy $A(Q_{o2}, K_{o2}) = 0$ and $E[F(\min\{Y, Q_{o2}\} + K_{o2})] = F(K_{c2})$.

Proof. When $\bar{Y} < Q_c$, we have from (3.5.13) that $A(\bar{Y}, 0) > 0$, which implies that $Q^*(0) = \bar{Y}$. Then, the proof of part (i) is identical to that of part (i) of Theorem 3.17 with Q_c replaced by $\bar{Y}$. When $c_r < c_{r8}$, we have $(d\tilde{\Pi}/dK)|_{K=0} > 0$ and consequently K_{o2} satisfies $E[F(S + K_{o2})] = F(K_{c2})$. When $A(\bar{Y}, K_{c2}) < 0$, in which case $K_1 > 0$, we have $K_{o2} \leq K_1$ if $(d\tilde{\Pi}/dK)|_{K=K_1} \leq 0$ and $K_{o2} > K_1$ otherwise. Then, parts (ii) and (iii) follow from part (iii) of Lemma 3.7 and the fact that $(d\tilde{\Pi}/dK)|_{K=K_1} \leq 0$ if $c_r \geq c_{r9}$. When $A(\bar{Y}, K_{c2}) \geq 0$ we have $c_{r9} = 0$ and part (ii) is a consequence of part (ii) of Lemma 3.7 (part (iii) is not applicable). □

3.5.3 Comparison of Models 1 and 2

In this section we compare the optimal order and reservation quantities for the two models. For a certain range of the reservation cost we show that when the option to buy from the backup supplier can be exercised after the demand becomes known (Model 2), it is optimal to order less from the primary supplier and reserve more with the backup supplier. To prove this we need to compare the critical reservation costs that determine whether the backup supplier should be used or not. To facilitate this comparison we obtain appropriate forms for the aforementioned critical values, denoted by $\tilde{c}_{r1}$ and $\tilde{c}_{r2}$ for Model 1 and 2, respectively. We have

$$\begin{aligned}\tilde{c}_{r1} &= (r+p-c_e)G(\min\{I,\bar{Q}\}) - (r+p-h)\int_0^{\min\{I,\bar{Q}\}} F(y)g(y)dy \\ &\quad + (c-c_e)^+\left[1 - G(\min\{I,\bar{Q}\})\right], \qquad (3.5.21)\\ \tilde{c}_{r2} &= (r+p-c_e)G(\bar{Q}) - (r+p-c_e)\int_0^{\bar{Q}} F(y)g(y)dy \\ &\quad + \frac{(r+p-c_e)(c-h)}{r+p-h}\left[1 - G(\bar{Q})\right]. \qquad (3.5.22)\end{aligned}$$

It is easy to verify that $\tilde{c}_{r1}$ reduces to c_{r1}, c_{r3}, c_{r4}, and c_{r6} for the special cases of Model 1, and $\tilde{c}_{r2}$ to c_{r7} and c_{r8} for Model 2.

In the following lemma we show that for Model 2 the backup supplier is used for a larger range of reservation cost values.

Lemma 3.9. $\tilde{c}_{r1} < \tilde{c}_{r2}$.

Proof. If $c_e < c$, in which case $I > Q_c \geq \bar{Q}$, we get

$$\tilde{c}_{r2} - \tilde{c}_{r1} = \frac{(r+p-c)(c_e-h)}{r+p-h}\left[1-G(\bar{Q})\right] + (c_e-h)\int_0^{\bar{Q}} F(y)g(y)dy > 0.$$

If $c_e \geq c$, we have $I < Q_c$. Then, if $\bar{Y} \leq I$, which implies that $\min\{I, \bar{Q}\} = \bar{Q} = \bar{Y}$, we get $\tilde{c}_{r2} - \tilde{c}_{r1} = (c_e - h)E[F(Y)] > 0$. On the other hand, if $\bar{Y} > I$, we have $\min\{I, \bar{Q}\} = I$ and

$$\begin{aligned} \tilde{c}_{r2} - \tilde{c}_{r1} &> (r+p-c_e)[G(\bar{Q}) - G(I)] - (r+p-c_e)\int_I^{\bar{Q}} F(y)g(y)dy \\ &= (r+p-c_e)\int_I^{\bar{Q}} [1-F(y)]g(y)dy \geq 0, \end{aligned}$$

completing the proof. □

The main result of the section is given in the following theorem.

Theorem 3.19. *For $c_r \geq \tilde{c}_{r1}$ we have $K_{o2} \geq K_{o1}$ and $Q_{o1} \geq Q_{o2}$.*

Proof. For $c_r \geq \tilde{c}_{r2}$ the theorem is satisfied with equalities because $K_{o1} = K_{o2} = 0$ and $Q_{o1} = Q_{o2} = \bar{Q}$. For $\tilde{c}_{r1} \leq c_r < \tilde{c}_{r2}$ we have $K_{o1} = 0$ and $Q_{o1} = \bar{Q}$. For Model 2, when $Q^*(K)$ is given by (3.5.17), we get from (3.5.20) that $dQ^*/dK < 0$, which combined with Theorems 3.17 and 3.18 yields $Q_{o2} \leq \bar{Q}$. □

Although we were not able to prove it, we believe that Theorem 3.19 holds in general. Our conjecture was supported by several numerical experiments with $c_r < \tilde{c}_{r1}$.

3.5.4 Effect of model parameters

In this section we discuss the effect of the cost and revenue parameters on the optimal order and reservation quantities. It can be shown that the optimal quantity associated with each supplier is decreasing with its cost and increasing with the cost of the other supplier. We omit the proof of this intuitive fact and focus on the effect of the rest of the parameters, r, p, and h, for which less intuitive results were obtained. First note that we do not need to consider r and p separately because the optimal solution depends on $v = r + p$. It is reasonable to expect that increasing values of v and h would give the retailer an incentive to order and reserve larger quantities because he would benefit more from their sale or salvage. Although this is true for the total quantity ordered and reserved, we have identified cases for which not both quantities are increasing, which is somewhat counterintuitive.

We present our results in the following sections, separately for each model. We restrict attention to probability distributions for which the maximum capacity is at least equal to the maximum demand, denoted by $\bar{X}$, which we believe is a realistic assumption. In this case we have $\bar{Y} > \min\{I, Q_c\}$ and $\bar{Y} > Q_c$ for Model 1 and 2, respectively, for any values of the problem parameters and the optimal solution is given by Theorems 3.12, 3.13, 3.15 and 3.17. We chose not to include the analysis for $\bar{Y} < \bar{X}$ because it is more complicated (Theorems 3.16 and 3.18 may also determine the optimal solution) and does not add any new insights.

3.5.4.1 Model 1

We start by proving a monotonicity property for the critical value of the reservation cost that determines whether the backup supplier should be used or not.

Lemma 3.10. *$\tilde{c}_{r1}$ is nondecreasing in v and h.*

Proof. Assuming $\bar{Y} > \min\{I, Q_c\}$ we have from (3.5.21)

$$\tilde{c}_{r1} = (c - c_e)^+ + (v - h) \int_0^{\min\{I, Q_c\}} [F(\min\{I, Q_c\}) - F(y)]g(y)dy. \tag{3.5.23}$$

Then, letting $J = \min\{I, Q_c\}$ we get

$$\begin{aligned} \frac{\partial \tilde{c}_{r1}}{\partial v} &= \int_0^J [F(J) - F(y)]g(y)dy + (v - h) \int_0^J \frac{\partial F(J)}{\partial v} g(y)dy, \\ \frac{\partial \tilde{c}_{r1}}{\partial h} &= -\int_0^J [F(J) - F(y)]g(y)dy + (v - h) \int_0^J \frac{\partial F(J)}{\partial h} g(y)dy. \end{aligned}$$

Both of the above are nonnegative because $\partial F(J)/\partial v = (\max\{c, c_e\} - h)/(v - h)^2$ and $\partial F(J)/\partial h = (v - \max\{c, c_e\})/(v - h)^2 = F(J)/(v - h)$. □

From (3.5.23) we get $\tilde{c}_{r1} < v - c_e$, so $\lim_{v \to c_r + c_e} \tilde{c}_{r1} < c_r$. It is also easy to show that $\lim_{v \to \infty} \tilde{c}_{r1} = \infty$. Therefore, there exists v_{r1} such that $\tilde{c}_{r1} > c_r$ for $v \geq v_{r1}$. Letting $H_{r1} = \lim_{h \to \min\{c, c_e\}} \tilde{c}_{r1}$, we have $\tilde{c}_{r1} < c_r$ for all $h < c, c_e$ if $H_{r1} \leq c_r$. Otherwise, there exists h_{r1} such that $\tilde{c}_{r1} > c_r$ for $h \geq h_{r1}$. In the following two theorems we obtain monotonicity properties for Q_{o1} and K_{o1} with respect to v and h for $v \in [v_{r1}, \infty)$ and $h \in [h_{r1}, \min\{c, c_e\})$, respectively. In any other case we have $Q_{o1} = Q_c$, which is increasing in v and h, and $K_{o1} = 0$.

Theorem 3.20. *If $c_e \geq c$, Q_{o1} and K_{o1} are increasing in v and h.*

Proof. The monotonicity property for Q_{o1} follows from the fact that $Q_{o1} = Q_c$ (Theorem 3.12). Turning to K_{o1}, we have from Theorem 3.13 that K_{o1} satisfies $(v - h) \int_0^{I - K_{o1}} [F(I) - F(y + K_{o1})]g(y)dy = c_r$. Differentiating both sides with respect to v and h and using

$\partial F(I)/\partial v = (1 - F(I))/(v - h)$ and $\partial F(I)/\partial h = F(I)/(v - h)$ we get

$$\frac{\partial K_{o1}}{\partial v}(v-h)\int_0^{I-K_{o1}} f(y+K_{o1})g(y)dy = (v-h)\frac{\partial F(I)}{\partial v}G(I-K_{o1})$$

$$+\int_0^{I-K_{o1}}[F(I)-F(y+K_{o1})]g(y)dy = \int_0^{I-K_{o1}}[1-F(y+K_{o1})]g(y)dy > 0, \tag{3.5.24}$$

$$\frac{\partial K_{o1}}{\partial h}(v-h)\int_0^{I-K_{o1}} f(y+K_{o1})g(y)dy = (v-h)\frac{\partial F(I)}{\partial h}G(I-K_{o1})$$

$$-\int_0^{I-K_{o1}}[F(I)-F(y+K_{o1})]g(y)dy = \int_0^{I-K_{o1}}F(y+K_{o1})g(y)dy > 0, \tag{3.5.25}$$

completing the proof. □

Theorem 3.21. *For $c_e < c$ we have*
i) K_{o1} is increasing in v and h,
ii) Q_{o1} is decreasing in v if f is nondecreasing, and increasing in h if f is nonincreasing.

Proof. According to Theorem 3.15, K_{o1} satisfies $(v-h)\int_0^{Q_c-K_{o1}}[F(Q_c)-F(y+K_{o1})]g(y)dy = c_r+c_e-c$, so that (3.5.24) and (3.5.25) hold with I replaced with Q_c, which proves part (i). To prove part (ii), we use $Q_{o1} = Q_c - K_{o1}$, $f(Q_c)(\partial Q_c/\partial v) = \partial F(Q_c)/\partial v = (1-F(Q_c))/(v-h)$, $f(Q_c)(\partial Q_c/\partial h) = \partial F(Q_c)/\partial h = F(Q_c)/(v-h)$, and (3.5.24) and (3.5.25) with Q_c instead of I to get

$$\frac{\partial Q_{o1}}{\partial v}(v-h)\int_0^{Q_{o1}} f(y+K_{o1})g(y)dy = [1-F(Q_c)]\frac{\int_0^{Q_{o1}} f(y+K_{o1})g(y)dy}{f(Q_c)}$$

$$-\int_0^{Q_{o1}}[1-F(y+K_{o1})]g(y)dy, \tag{3.5.26}$$

$$\frac{\partial Q_{o1}}{\partial h}(v-h)\int_0^{Q_{o1}} f(y+K_{o1})g(y)dy = F(Q_c)\frac{\int_0^{Q_{o1}} f(y+K_{o1})g(y)dy}{f(Q_c)}$$

$$-\int_0^{Q_{o1}} F(y+K_{o1})g(y)dy. \tag{3.5.27}$$

Noting that $y + K_{o1} < Q_c$ for $0 < y < Q_{o1}$, it is easy to see that the righthand side of (3.5.26) is negative if f is nondecreasing and the righthand side of (3.5.27) is positive if f is nonincreasing. □

We end the section with the outcome of a numerical study on the behavior of the optimal order quantity to the primary supplier when the assumptions in part (ii) of Theorem 3.21

are relaxed. Specifically, we examined the dependence of Q_{o1} on v (resp. h) for several examples with demand that follows a triangular distribution with decreasing (resp. increasing) probability density function. In all of these examples Q_{o1} exhibited the monotonic behavior described in the theorem. Table 3.1 summarizes the findings of this section.

Table 3.1: Effect of model parameters on optimal order and reservation quantities

Parameters	Q_o	K_o
$c\uparrow$	$\downarrow$	$\uparrow$
$c_e, c_r\uparrow$	$\uparrow$	$\downarrow$
$r, p\uparrow$	$\downarrow$	$\uparrow$
$h\uparrow$	$\uparrow$	$\uparrow$

3.5.4.2 Model 2

As with Model 1, we first show that the critical value of the reservation cost varies monotonically with v and h.

Lemma 3.11. *$\tilde{c}_{r2}$ is increasing in v and decreasing in h.*

Proof. For $\bar{Y} > Q_c$ we have from (3.5.22)

$$\tilde{c}_{r2} = c_{r7} = (v - c_e)\left[1 - \int_0^{Q_c} F(y)g(y)dy - F(Q_c)[1 - G(Q_c)]\right], \tag{3.5.28}$$

which yields

$$\begin{aligned}
\frac{\partial \tilde{c}_{r2}}{\partial v} &= 1 - \int_0^{Q_c} F(y)g(y)dy - F(Q_c)[1 - G(Q_c)] \\
&\quad -(v - c_e)\frac{\partial F(Q_c)}{\partial v}[1 - G(Q_c)] \\
&\geq 1 - F(Q_c) - \frac{(v - c_e)(c - h)}{(v - h)^2}[1 - G(Q_c)] \\
&= \frac{c - h}{v - h}\left[1 - \frac{v - c_e}{v - h}[1 - G(Q_c)]\right] > 0, \\
\frac{\partial \tilde{c}_{r2}}{\partial h} &= -(v - c_e)\frac{\partial F(Q_c)}{\partial h}[1 - G(Q_c)] \\
&= -\frac{v - c_e}{v - h}F(Q_c)[1 - G(Q_c)] < 0,
\end{aligned}$$

where we have used the expressions derived in the proof of Lemma 3.10 for $\partial F(Q_c)/\partial v$ and $\partial F(Q_c)/\partial h$. □

From (3.5.28) we get $\tilde{c}_{r2} < v - c_e$ and $\lim_{v\to\infty} \tilde{c}_{r2} = \infty$, so that there exists v_{r2} such that $\tilde{c}_{r2} > c_r$ for $v \geq v_{r2}$. With H_{r2} denoting the value of $\tilde{c}_{r2}$ for $h = 0$, we have $\tilde{c}_{r2} < c_r$ for all $h < c, c_e$ if $H_{r2} \leq c_r$. Otherwise, there exists $h_{r2} \leq \min\{c, c_e\}$ such that $\tilde{c}_{r2} > c_r$ for $h < h_{r2}$. In the following theorem we obtain monotonicity properties for Q_{o2} and K_{o2} with respect to v and h for $v \in [v_{r2}, \infty)$ and $h \in [0, h_{r2})$, respectively. In any other case we have $Q_{o2} = Q_c$, which is increasing in v and h, and $K_{o2} = 0$.

Theorem 3.22. *i) K_{o2} is increasing in v and nonincreasing in h.*
ii) Q_{o2} is nonincreasing in v if f is nondecreasing, and increasing in h.

Proof. Differentiating with respect to v and h both sides of the equations that determine the optimal order and reservation quantities (part (ii) of Theorem 3.17) we get

$$C \cdot \begin{bmatrix} \frac{\partial Q_{o2}}{\partial h} \\ \frac{\partial K_{o2}}{\partial h} \end{bmatrix} = \begin{bmatrix} F(Q_{o2}) \\ 0 \end{bmatrix}, \tag{3.5.29}$$

$$C \cdot \begin{bmatrix} \frac{\partial Q_{o2}}{\partial v} \\ \frac{\partial K_{o2}}{\partial v} \end{bmatrix} = \begin{bmatrix} 1 - F(Q_{o2} + K_{o2}) \\ (v - c_e)^{-1}[1 - E[F(\min\{Y, Q_{o2}\} + K_{o2})] \end{bmatrix}, \tag{3.5.30}$$

where

$$C = \begin{bmatrix} (v - c_e)f(Q_{o2} + K_{o2}) + (c_e - h)f(Q_{o2}) & (v - c_e)f(Q_{o2} + K_{o2}) \\ f(Q_{o2} + K_{o2})[1 - G(Q_{o2})] & E\left[f(\min\{Y, Q_{o2}\} + K_{o2})\right] \end{bmatrix}. \tag{3.5.31}$$

Solving the linear systems of two equations with two unknowns defined by (3.5.29)-(3.5.31) we get $\partial K_{o2}/\partial h = A_h/\Delta$, $\partial Q_{o2}/\partial h = B_h/\Delta$, $\partial K_{o2}/\partial v = A_v/\Delta$, and $\partial Q_{o2}/\partial v = B_v/\Delta$, where

$$\begin{aligned}
\Delta &= (v - c_e)f(Q_{o2} + K_{o2}) \int_0^{Q_{o2}} f(y + K_{o2})g(y)dy \\
&\quad + (c_e - h)f\left(Q_{o2}\right) E\left[f(\min\{Y, Q_{o2}\} + K_{o2})\right], \\
A_h &= -F(Q_{o2})f(Q_{o2} + K_{o2})[1 - G(Q_{o2})], \\
B_h &= F(Q_{o2})E\left[f(\min\{Y, Q_{o2}\} + K_{o2})\right], \\
A_v &= \frac{(c_e - h)}{(v - c_e)} f(Q_{o2})[1 - E[F(\min\{Y, Q_{o2}\} + K_{o2})] \\
&\quad + f(Q_{o2} + K_{o2}) \int_0^{Q_{o2}} [1 - F(y + K_{o2})]g(y)dy, \\
B_v &= [1 - F(Q_{o2} + K_{o2})]E\left[f(\min\{Y, Q_{o2}\} + K_{o2})\right] \\
&\quad - f(Q_{o2} + K_{o2})[1 - E[F(\min\{Y, Q_{o2}\} + K_{o2})].
\end{aligned}$$

It is easy to see that Δ, B_h, and A_v are strictly positive. On the other hand, A_h and B_v are equal to zero if $Q_{o2} + K_{o2} \geq \bar{X}$. Otherwise, A_h is negative and if f is nondecreasing, B_v is also negative. □

Finally, we conducted a numerical investigation assuming that demand is triangularly distributed with decreasing pdf, and thus relaxing the assumption in part (ii) of Theorem 3.22. In all the problem instances we considered we observed that Q_{o2} was decreasing in v. The results of this section are summarized in Table 3.2.

Table 3.2: Effect of model parameters on optimal order and reservation quantities

Parameters	Q_o	K_o
$c \uparrow$	$\downarrow$	$\uparrow$
$c_e, c_r \uparrow$	$\uparrow$	$\downarrow$
$r, p \uparrow$	$\downarrow$	$\uparrow$
$h \uparrow$	$\uparrow$	$\downarrow$

3.6 Conclusions

In this chapter we studied a special class of newsvendor models with two types of suppliers: primary suppliers who are subject to random yield or random capacity and a reliable supplier who can be used as backup after the delivery from primary suppliers and before or after the demand becomes known as well, provided that his capacity has been reserved in advance. We derived expressions for the order and reservation quantities that maximize the expected profit of the retailer, from which we obtained several interesting insights. One such insight had to do with situations where reservation is irrelevant in the sense that it is certain that all of the reserved quantity will be eventually bought. This may happen when the option to buy from the backup supplier is exercised before the demand becomes known and the backup supplier is cheaper than the primary supplier after excluding reservation costs. Another rather surprising fact is that increased sales and salvage prices do not necessarily result in higher order and reservation quantities. Finally, as a natural continuation of this work, we propose the study of multi-period models as a topic for future research.

Appendix A

Proof of Lemma 2.1

First, we show $V(x_1, x_2) > V(x_1, x_2 - 1)$ for $x_1 \geq 0$, $x_2 \geq 1$. We use induction on x_1. For $x_1 = 0$, the result follows from (2.3.5) and (2.3.6). For $x_1 \geq 1$, if $(0, \rho_2^*)$ is the optimal allocation at (x_1, x_2), we have from (2.3.1) and (2.3.2)

$$V(x_1, x_2) - V(x_1, x_2 - 1) = \frac{h_1 x_1 + h_2 x_2}{\rho_2^*} > 0.$$

Otherwise, (2.3.1) yields

$$V(x_1, x_2) - V(x_1, x_2 - 1) = h_2 + \min_{(\rho_1, \rho_2) \in A(x_1, x_2)} W_{\rho_1, \rho_2}(x_1, x_2) - \min_{(\rho_1, \rho_2) \in A(x_1, x_2 - 1)} W_{\rho_1, \rho_2}(x_1, x_2 - 1).$$

Let $(\rho_1^*, \rho_2^*) = \arg\min_{(\rho_1, \rho_2) \in A(x_1, x_2)} W_{\rho_1, \rho_2}(x_1, x_2)$. Then, because $(\rho_1^*, 0) \in A(x_1, x_2 - 1)$, we have

$$V(x_1, x_2) - V(x_1, x_2 - 1) \geq h_2 + W_{\rho_1^*, \rho_2^*}(x_1, x_2) - W_{\rho_1^*, 0}(x_1, x_2 - 1),$$

which combined with (2.3.2) leads to

$$V(x_1, x_2) - V(x_1, x_2 - 1) \geq \frac{h_2 + \rho_1^* \left[V(x_1 - 1, x_2 + 1) - V(x_1 - 1, x_2)\right]}{\rho_1^* + \rho_2^*} > 0,$$

where the last inequality follows from the induction hypothesis.

We now show that $V(x_1, x_2) > V(x_1 - 1, x_2)$ for $x_1 \geq 1$, $x_2 \geq 0$. The proof is by induction on x_2. The induction base is established by noting that for all $x_1 \geq 1$ we have $V(x_1, 0) > V(x_1 - 1, 1) > V(x_1 - 1, 0)$, where the first inequality follows from (2.3.3) and (2.3.4) and the second from the monotonicity of the value function with respect to its second argument. For $x_2 \geq 1$, if $(\rho_1^*, 0)$ is the optimal allocation at (x_1, x_2), we have

$$V(x_1, x_2) - V(x_1 - 1, x_2) > V(x_1, x_2) - V(x_1 - 1, x_2 + 1) = \frac{h_1 x_1 + h_2 x_2}{\rho_1^*} > 0. \qquad \text{(A.1)}$$

Otherwise,

$$V(x_1, x_2) - V(x_1 - 1, x_2) \geq h_1 + W_{\rho_1^*, \rho_2^*}(x_1, x_2) - W_{0, \rho_2^*}(x_1 - 1, x_2),$$

where $(\rho_1^*, \rho_2^*) = \arg\min_{(\rho_1, \rho_2) \in A(x_1, x_2)} W_{\rho_1, \rho_2}(x_1, x_2)$. Applying (2.3.2) we get

$$\begin{aligned}(\rho_1^* + \rho_2^*)\left[V(x_1, x_2) - V(x_1 - 1, x_2)\right] \;\; &\geq \;\; h_1 \\ &+ \rho_1^* \left[V(x_1 - 1, x_2 + 1) - V(x_1 - 1, x_2)\right] + \rho_2^* \left[V(x_1, x_2 - 1) - V(x_1 - 1, x_2 - 1)\right],\end{aligned}$$

which is positive because of the monotonicity of the value function with respect to x_2 and the induction hypothesis.

Proof of Lemma 2.3

The proof is by induction on x_2 and is similar to that of Lemma 2.1. For $x_2 = 0$, the result follows from (2.3.3) and (2.3.4). For $x_2 \geq 1$, if $(\rho_1^*, 0)$ is the optimal allocation at (x_1, x_2), $f(x_1, x_2) > 0$ by (A.1). Otherwise,

$$\begin{aligned} f(x_1,x_2) &= V(x_1,x_2) - V(x_1-1,x_2+1) \geq h_1 - h_2 + W_{\rho_1^*,\rho_2^*}(x_1,x_2) - W_{0,\rho_2^*}(x_1-1,x_2+1) \\ &= h_1 - h_2 + \rho_2^* f(x_1,x_2-1) + (1-\rho_1^*-\rho_2^*) f(x_1,x_2), \end{aligned}$$

where $(\rho_1^*, \rho_2^*) = \arg\min_{(\rho_1,\rho_2)\in A(x_1,x_2)} W_{\rho_1,\rho_2}(x_1,x_2)$ and the last equality follows from (2.3.2). Then $f(x_1, x_2) > 0$ because $h_1 \geq h_2$ and $f(x_1, x_2 - 1) > 0$ (induction hypothesis).

Proof of Lemma 2.4

We start by deriving a recursive equation for $\tilde{d}(1, x_2)$, $x_2 \geq 0$. To do that we use optimality equations (2.3.3) and (2.3.5)-(2.3.8) and, where applicable, identities $\min(a, b) = a + (b - a)^-$ and $\min(a, b) = b - (b - a)^+$ for the first and second terms, respectively, of the differences appearing in the definition of $f(1, x_2)$ and $g(1, x_2)$. We get

$$\begin{aligned} \tilde{d}(1,0) &= \xi_1(h_1-h_2) + \xi_1(\nu_2+\xi_2)V(0,1) + (\nu_2+\mu_1+\mu_2+\xi_2)\tilde{d}(1,0), \quad (A.2) \\ \tilde{d}(1,1) &= \xi_1(h_1-h_2) - \mu_2 h_2 + \nu_2\tilde{d}(1,0) + \mu_1\tilde{d}(1,1) + \nu_1\mu_2 g(0,2) \\ &\quad +\xi_1\xi_2 f(1,1) + \mu_2^2 g(1,1) + \xi_1\bar{d}(1,1)^- + \mu_2\bar{d}(1,1)^+, \quad (A.3) \\ \tilde{d}(1,x_2) &= \xi_1(h_1-h_2) - \mu_2 h_2 + \nu_2\tilde{d}(1,x_2-1) + (\mu_1+\xi_2)\tilde{d}(1,x_2) \\ &\quad +\nu_1\mu_2 g(0,x_2+1) + \xi_1\tilde{d}(1,x_2)^- + \mu_2\tilde{d}(1,x_2)^+ \\ &\quad +\mu_2\left[\bar{d}(1,1)^-\mathbf{1}(x_2=2) + \tilde{d}(1,x_2-1)^-\mathbf{1}(x_2>2)\right], \quad x_2 \geq 2. \quad (A.4) \end{aligned}$$

For $x_2 \geq 2$ we get from (2.3.6) that $g(0, x_2) = -h_2x_2/(\mu_2 + \nu_2)$ and consequently

$$g(0,x_2) - g(0,x_2+1) = \frac{h_2}{\mu_2+\nu_2}. \quad (A.5)$$

To prove $\tilde{d}(1, 0) > \tilde{d}(1, 1)$ we consider two cases for $\bar{d}(1, 1)$. When $\bar{d}(1, 1) < 0$ we get from (A.3)

$$\begin{aligned} \tilde{d}(1,1) &= \xi_1(h_1-h_2) - \mu_2 h_2 + \nu_2\tilde{d}(1,0) + (\mu_1+\mu_2)\tilde{d}(1,1) + \nu_1\mu_2 g(0,2) \\ &\quad +\xi_1(\xi_2-\mu_2)f(1,1) + \xi_1\bar{d}(1,1) \\ &< \xi_1(h_1-h_2) - \mu_2 h_2 + \nu_2\tilde{d}(1,0) + (\mu_1+\mu_2)\tilde{d}(1,1) \quad (A.6) \end{aligned}$$

because $g(0, 2) < 0$, $\mu_2 > \xi_2$, $f(1, 1) > 0$ and $\bar{d}(1, 1) < 0$. Then (A.2) and (A.6) yield

$$\tilde{d}(1,0) - \tilde{d}(1,1) > \frac{\mu_2 h_2 + \xi_1(\nu_2+\xi_2)V(0,1) + \xi_2\tilde{d}(1,0)}{1-\mu_1-\mu_2} > 0$$

because $\tilde{d}(1, 0) = \xi_1 f(1, 0) > 0$. When $\bar{d}(1, 1) \geq 0$, substituting $\bar{d}(1, 1)^+ = \bar{d}(1, 1) = \xi_1 f(1, 1) + \xi_2 g(1, 1)$ in (A.3) we get

$$\tilde{d}(1,1) = \xi_1(h_1-h_2) - \mu_2 h_2 + \nu_2\tilde{d}(1,0) + (\mu_1+\mu_2+\xi_2)\tilde{d}(1,1) + \nu_1\mu_2 g(0,2),$$

which combined with (A.2) gives

$$\tilde{d}(1,0)-\tilde{d}(1,1)=\frac{\mu_2 h_2+\xi_1(\nu_2+\xi_2)V(0,1)-\nu_1\mu_2 g(0,2)}{1-\mu_1-\mu_2-\xi_2}>0.$$

Next we show $\tilde{d}(1,1)>\tilde{d}(1,2)$. Using $\bar{d}(1,1)=\bar{d}(1,1)^-+\bar{d}(1,1)^+$ in (A.3) we get

$$\begin{aligned}\tilde{d}(1,1) &= \xi_1(h_1-h_2)-\mu_2 h_2+\nu_2\tilde{d}(1,0)+(\mu_1+\xi_2)\tilde{d}(1,1)+\nu_1\mu_2 g(0,2)\\ &\quad +\mu_2\tilde{d}(1,1)+\xi_1\bar{d}(1,1)^- -\mu_2\bar{d}(1,1)^-. \end{aligned}\tag{A.7}$$

We consider two cases for $\tilde{d}(1,1)$. Assume first that $\tilde{d}(1,1)<0$. Because $\tilde{d}(1,1)<\bar{d}(1,1)$ and $\tilde{d}(1,1)^+=0$, we get from (A.7) and (A.4)

$$\begin{aligned}\tilde{d}(1,1)-\tilde{d}(1,2) &> \nu_2\tilde{d}(1,0)+(\mu_2-\nu_2)\tilde{d}(1,1)+(\mu_1+\xi_2)\left[\tilde{d}(1,1)-\tilde{d}(1,2)\right]\\ &\quad +\mu_1\nu_2\left[g(0,2)-g(0,3)\right]+\xi_1\left[\tilde{d}(1,1)^- -\tilde{d}(1,2)^-\right]+\mu_2\left[\bar{d}(1,1)^+ -\tilde{d}(1,2)^+\right],\end{aligned}$$

and $\tilde{d}(1,1)-\tilde{d}(1,2)>0$ follows from $\mu_2\le\nu_2$, (A.5) and Lemma 2.2. When $\tilde{d}(1,1)\ge 0$, which implies that $\bar{d}(1,1)\ge 0$ as well, (A.7) and (A.4) yield

$$\begin{aligned}\tilde{d}(1,1)-\tilde{d}(1,2) &= \nu_2\left[\tilde{d}(1,0)-\tilde{d}(1,1)\right]+(\mu_1+\xi_2)\left[\tilde{d}(1,1)-\tilde{d}(1,2)\right]\\ &\quad +\mu_1\nu_2\left[g(0,2)-g(0,3)\right]+\mu_2\left[\bar{d}(1,1)^+ -\tilde{d}(1,2)^+\right]-\xi_1\tilde{d}(1,2)^-,\end{aligned}$$

and $\tilde{d}(1,1)-\tilde{d}(1,2)>0$ follows from $\tilde{d}(1,0)>\tilde{d}(1,1)$, (A.5) and Lemma 2.2. From (A.4) we get

$$\begin{aligned}\tilde{d}(1,2)-\tilde{d}(1,3) &= \nu_2\left[\tilde{d}(1,1)-\tilde{d}(1,2)\right]+(\mu_1+\xi_2)\left[\tilde{d}(1,2)-\tilde{d}(1,3)\right]\\ &\quad +\mu_1\nu_2\left[g(0,3)-g(0,4)\right]+\mu_2\left[\bar{d}(1,1)^- -\tilde{d}(1,2)^-\right]\\ &\quad +\xi_1\left[\tilde{d}(1,2)^- -\tilde{d}(1,3)^-\right]+\mu_2\left[\tilde{d}(1,2)^+ -\tilde{d}(1,3)^+\right],\end{aligned}$$

and $\tilde{d}(1,2)-\tilde{d}(1,3)>0$ follows from $\bar{d}(1,1)>\tilde{d}(1,1)>\tilde{d}(1,2)$, (A.5) and Lemma 2.2. For $x_2\ge 3$, $\tilde{d}(1,x_2)-\tilde{d}(1,x_2+1)>0$ can be proved by a straightforward induction on x_2 based on (A.4) and application of Lemma 2.2, thus completing the proof of the first part of the lemma.

We now turn to the proof of part (ii). Because $\tilde{d}(1,x_2)$ is a decreasing sequence, its limit as $x_2\to\infty$ exists. Moreover, it is easy to show that $\lim_{x_2\to\infty}\tilde{d}(1,x_2)^-=\left[\lim_{x_2\to\infty}\tilde{d}(1,x_2)\right]^-$ and $\lim_{x_2\to\infty}\tilde{d}(1,x_2)^+=\left[\lim_{x_2\to\infty}\tilde{d}(1,x_2)\right]^+$. Assume $\tilde{L}=\lim_{x_2\to\infty}\tilde{d}(1,x_2)$ is finite. By taking limits in (A.4) we get

$$\tilde{L}=\xi_1(h_1-h_2)-\mu_2 h_2+(\nu_2+\mu_1+\xi_2)\tilde{L}+\nu_1\mu_2\lim_{x_2\to\infty}g(0,x_2)+\xi_1\left(\tilde{L}-\tilde{L}^+\right)+\mu_2\left(\tilde{L}^+ +\tilde{L}^-\right),$$

leading to

$$\nu_1\tilde{L}=\xi_1(h_1-h_2)-\mu_2 h_2-\xi_1\tilde{L}^+ +\nu_1\mu_2\lim_{x_2\to\infty}g(0,x_2),$$

which is a contradiction because $g(0,x_2)=-h_2x_2/(\nu_2+\mu_2)$ tends to $-\infty$ as $x_2\to\infty$.

Proof of Lemma 2.5

The proof is by induction on x_1. We start with $x_1 = 1$ to establish the induction base. We derive the following recursive equation for $d(1,x_2)$ in the same way we did for $\tilde{d}(1,x_2)$.

$$\begin{aligned}
d(1,0) &= \mu_1(h_1-h_2)+\mu_1(\nu_2+\xi_2)V(0,1)+(\nu_2+\mu_1+\mu_2+\xi_2)d(1,0), && \text{(A.8)}\\
d(1,1) &= \mu_1(h_1-h_2)-\mu_2h_2+\nu_2 d(1,0)+(\mu_1+\mu_2+\xi_2)d(1,1)\\
&\quad +\mu_2(\nu_1+\xi_1-\mu_1)g(0,2)+(\mu_1-\mu_2)\bar{d}(1,1)^-, && \text{(A.9)}\\
d(1,x_2) &= \mu_1(h_1-h_2)-\mu_2h_2+\nu_2 d(1,x_2-1)+(\mu_1+\mu_2+\xi_2)d(1,x_2)\\
&\quad +\mu_2(\nu_1+\xi_1-\mu_1)g(0,x_2+1)+(\mu_1-\mu_2)\tilde{d}(1,x_2)^-\\
&\quad +\mu_2\left[\bar{d}(1,1)^-\mathbf{1}(x_2=2)+\tilde{d}(1,x_2-1)^-\mathbf{1}(x_2>2)\right], \quad x_2\ge 2. && \text{(A.10)}
\end{aligned}$$

Then $d(1,x_2)-d(1,x_2+1)>0$ can be proved by induction on x_2 based on (A.8)-(A.10), using the facts that $\nu_1+\xi_1>\mu_1\ge\mu_2$, $g(0,x_2)$ is negative and decreasing, $\bar{d}(1,1)>\tilde{d}(1,1)$, and $\tilde{d}(1,x_2)$ is decreasing. For $x_1\ge 2$ we have

$$\begin{aligned}
d(x_1,0) &= \mu_1(h_1-h_2)+\nu_1\mu_1 f(x_1-1,1)-\mu_1(\nu_2+\xi_2)g(x_1-1,1)\\
&\quad +(\nu_2+\mu_2+\xi_1+\xi_2)d(x_1,0)\\
&\quad +\mu_1\left[\bar{d}(1,1)^+\mathbf{1}(x_1=2)+\hat{d}(x_1-1,1)^+\mathbf{1}(x_1>2)\right], && \text{(A.11)}\\
d(x_1,1) &= \mu_1(h_1-h_2)-\mu_2h_2+\nu_1 d(x_1-1,2)+\nu_2 d(x_1,0)\\
&\quad +(\xi_1+\xi_2)d(x_1,1)+\mu_2(\mu_2-\xi_2)g(x_1,1)+\mu_1\hat{d}(x_1,1)^-+\mu_2\hat{d}(x_1,1)^+\\
&\quad +\mu_1\left[\tilde{d}(1,2)^+\mathbf{1}(x_1=2)+d(x_1-1,2)^+\mathbf{1}(x_1>2)\right], && \text{(A.12)}\\
d(x_1,x_2) &= \mu_1(h_1-h_2)-\mu_2h_2+\nu_1 d(x_1-1,x_2+1)+\nu_2 d(x_1,x_2-1)\\
&\quad +(\xi_1+\xi_2)d(x_1,x_2)+\mu_1 d(x_1,x_2)^-+\mu_2 d(x_1,x_2)^+\\
&\quad +\mu_1\left[\tilde{d}(1,x_2+1)^+\mathbf{1}(x_1=2)+d(x_1-1,x_2+1)^+\mathbf{1}(x_1>2)\right]\\
&\quad +\mu_2\left[\hat{d}(x_1,1)^-\mathbf{1}(x_2=2)+d(x_1,x_2-1)^-\mathbf{1}(x_2>2)\right], \quad x_2\ge 2. && \text{(A.13)}
\end{aligned}$$

To show $d(x_1,0)>d(x_1,1)>d(x_1,2)$ we consider two cases for $\hat{d}(x_1,1)$. First, assume $\hat{d}(x_1,1)<0$ which implies $d(x_1,1)<0$ as well. Then, $d(x_1,1)<d(x_1,0)$ because $d(x_1,0)=\mu_1 f(x_1,0)>0$. Noting that $(\mu_2-\xi_2)g(x_1,1)=d(x_1,1)-\hat{d}(x_1,1)$, $\hat{d}(x_1,1)>d(x_1,1)$, and $\hat{d}(x_1,1)^+=d(x_1,1)^+=0$, we get from (A.12) and (A.13)

$$\begin{aligned}
d(x_1,1)-d(x_1,2) &> \nu_2 d(x_1,0)+\nu_1\left[d(x_1-1,2)-d(x_1-1,3)\right]+\\
&\quad (\mu_2-\nu_2)d(x_1,1)+\mu_1\left[\tilde{d}(1,2)^+-\tilde{d}(1,3^+)\right]\mathbf{1}(x_1=2)\\
&\quad +\mu_1\left[d(x_1-1,2)^+-d(x_1-1,3)^+\right]\mathbf{1}(x_1>2)+(\xi_1+\xi_2)\left[d(x_1,1)-d(x_1,2)\right]\\
&\quad +\mu_1\left[d(x_1,1)^--d(x_1,2)^-\right]+\mu_2\left[d(x_1,1)^+-d(x_1,2)^+\right]. && \text{(A.14)}
\end{aligned}$$

Because $d(x_1,0)>0$, $\mu_2\le\nu_2$, $d(x_1,1)<0$ and $\tilde{d}(1,x_2)$ is decreasing, we get $d(x_1,1)>d(x_1,2)$ by applying the induction hypothesis and Lemma 2.2. When $\hat{d}(x_1,1)\ge 0$, we have in (A.12) $\mu_2(\mu_2-\xi_2)g(x_1,1)+\mu_2\hat{d}(x_1,1)^+=\mu_2 d(x_1,1)$. Therefore, taking also into account that $\mu_1 f(x_1-1,1)>d(x_1-1,1)$ and $g(x_1-1,1)<0$, we obtain from (A.11) and (A.12)

$$\begin{aligned}(1-\xi_1-\xi_2-\mu_2)\left[d(x_1,0)-d(x_1,1)\right] &> \mu_2h_2+\nu_1\left[d(x_1-1,1)-d(x_1-1,2)\right]+\\ &+\mu_1\left[\bar{d}(1,1)^+-\tilde{d}(1,2)^+\right]\mathbf{1}(x_1=2)+\mu_1\left[\hat{d}(x_1-1,1)^+-d(x_1-1,2)^+\right]\mathbf{1}(x_1>2).\end{aligned}$$

The righthand side of the equation above is positive by the induction hypothesis, $\bar{d}(1,1)>\tilde{d}(1,1)>\tilde{d}(1,2)$, and $\hat{d}(x_1-1,1)>d(x_1-1,1)$ for $x_1>2$. When $d(x_1,1)<0$, (A.12) and (A.13) yield (A.14) without the second to last term, so $d(x_1,1)>d(x_1,2)$ is proved similarly. When $d(x_1,1)\geq 0$, in which case $d(x_1,1)=d(x_1,1)^+$, we get

$$\begin{aligned}d(x_1,1)-d(x_1,2) &> \nu_1\left[d(x_1-1,2)-d(x_1-1,3)\right]+\nu_2\left[d(x_1,0)-d(x_1,1)\right]\\ &+\mu_1\left[\tilde{d}(1,2)^+-\tilde{d}(1,3)^+\right]\mathbf{1}(x_1=2)+\mu_1\left[d(x_1-1,2)^+-d(x_1-1,3)^+\right]\mathbf{1}(x_1>2)\\ &+(\xi_1+\xi_2)\left[d(x_1,1)-d(x_1,2)\right]+\mu_2\left[d(x_1,1)^+-d(x_1,2)^+\right]\end{aligned}$$

and $d(x_1,1)>d(x_1,2)$ follows from $d(x_1,0)>d(x_1,1)$, $\tilde{d}(1,x_2)$ being decreasing, the induction hypothesis and Lemma 2.2. Next, (A.13) yields

$$\begin{aligned}d(x_1,2)-d(x_1,3) &= \nu_1\left[d(x_1-1,3)-d(x_1-1,4)\right]+\nu_2\left[d(x_1,1)-d(x_1,2)\right]\\ &+\mu_1\left[\tilde{d}(1,3)^+-\tilde{d}(1,4)^+\right]\mathbf{1}(x_1=2)+\mu_1\left[d(x_1-1,3)^+-d(x_1-1,4)^+\right]\mathbf{1}(x_1>2)\\ &+\mu_2\left[\hat{d}(x_1,1)^--d(x_1,2)^-\right]+(\xi_1+\xi_2)\left[d(x_1,2)-d(x_1,3)\right]\\ &+\mu_1\left[d(x_1,2)^--d(x_1,3)^-\right]+\mu_2\left[d(x_1,2)^+-d(x_1,3)^+\right]\end{aligned}$$

and $d(x_1,2)>d(x_1,3)$ follows from $d(x_1,1)>d(x_1,2)$, $\tilde{d}(1,x_2)$ being decreasing, $\hat{d}(x_1,1)>d(x_1,1)$, the induction hypothesis and Lemma 2.2. Finally, for $x_2\geq 3$, $d(x_1,x_2)-d(x_1,x_2+1)>0$ is proved easily by induction on x_2 based on (A.13) and application of the induction hypothesis for x_1 and Lemma 2.2.

The second part of the lemma is proved by induction on x_1. Let $L(x_1)=\lim_{x_2\to\infty}d(x_1,x_2)$; this limit exists because $d(x_1,x_2)$ is decreasing in x_2. From (A.10) we have

$$\begin{aligned}d(1,x_2) &\leq \mu_1(h_1-h_2)-\mu_2h_2+\nu_2 d(1,x_2-1)+(\mu_1+\mu_2+\xi_2)d(1,x_2)\\ &+\mu_2(\nu_1+\xi_1-\mu_1)g(0,x_2+1).\end{aligned}$$

Assuming $L(1)$ is finite and taking limits on both sides we get

$$(\nu_1+\xi_1)L(1)\leq \mu_1(h_1-h_2)-\mu_2h_2+\mu_2(\nu_1+\xi_1-\mu_1)\lim_{x_2\to\infty}g(0,x_2)=-\infty,$$

clearly a contradiction. Let now $L(x_1-1)=-\infty$ be the induction hypothesis. Taking also into account that $\tilde{L}=-\infty$, we get from (A.13) for x_2 sufficiently large

$$\begin{aligned}d(x_1,x_2) &= \mu_1(h_1-h_2)-\mu_2h_2+\nu_1 d(x_1-1,x_2+1)+\nu_2 d(x_1,x_2-1)\\ &+(\xi_1+\xi_2)d(x_1,x_2)+\mu_1 d(x_1,x_2)^-+\mu_2 d(x_1,x_2)^++\mu_2 d(x_1,x_2-1)^-.\end{aligned}$$

Assuming $L(x_1)$ is finite and taking limits on both sides we get

$$\nu_1 L(x_1)=\mu_1(h_1-h_2)-\mu_2h_2-\mu_1 L(x_1)^++\nu_1 L(x_1-1)=-\infty,$$

which is a contradiction. Therefore, $L(x_1)=-\infty$, $x_1\geq 1$, completing the proof of the lemma.

Proof of Lemma 2.6

The proof of part (i) is by induction on x_2. Note that $d(2,1) \geq 0$ implies $\hat{d}(2,1) \geq 0$ and for $x_2 \geq 2$, $d(2,x_2) \geq 0$ implies $d(2,x_2-1) \geq 0$ by Lemma 2.5(i). Taking the above into account and after some straightforward algebraic manipulations we get from (A.12), (A.13) and (A.10)

$$d(2,x_2) \geq \mu_1(h_1-h_2) - \mu_2 h_2 + \nu_1 d(1,x_2+1) + \nu_2 d(2,x_2-1) + (\xi_1+\xi_2+\mu_2)d(2,x_2),$$
$$d(1,x_2+1) \leq \mu_1(h_1-h_2) - \mu_2 h_2 + \nu_2 d(1,x_2) + (\mu_1+\mu_2+\xi_2)d(1,x_2+1),$$

from which we obtain

$$d(2,x_2) - d(1,x_2+1) \geq \frac{\nu_2\left[d(2,x_2-1)-d(1,x_2)\right] + (\nu_1+\xi_1-\mu_1)d(1,x_2+1)}{\nu_1+\nu_2+\mu_1}. \qquad \text{(A.15)}$$

We assume that $d(1,x_2+1) \geq 0$ because otherwise there is nothing to prove. Then, $d(2,1) \geq d(1,2)$ follows from $d(2,0) > d(1,0) > d(1,1)$ and (A.15) for $x_2 = 1$, establishing the induction base. For $x_2 \geq 2$, because $d(2,x_2-1) \geq 0$, the induction hypothesis implies that $d(2,x_2-1) - d(1,x_2) \geq 0$, so we get $d(2,x_2) - d(1,x_2+1) \geq 0$ from (A.15).

Before proceeding to parts (ii) and (iii), we use the optimality equations to get for $x_1 \geq 2$

$$\hat{d}(x_1,1) = C(x_1) + (\xi_1+\mu_2)\hat{d}(x_1,1) + \mu_1\hat{d}(x_1,1)^- + \xi_2\hat{d}(x_1,1)^+, \qquad \text{(A.16)}$$

where

$$\begin{aligned} C(x_1) &= \mu_1(h_1-h_2) - \xi_2 h_2 + \nu_1\left[\mu_1 f(x_1-1,2) + \xi_2 g(x_1-1,2)\right] + \nu_2 d(x_1,0) \\ &\quad -\mu_1(\mu_2-\xi_2)g(x_1-1,2) + \mu_1\left[\tilde{d}(1,2)^+\mathbf{1}(x_1=2) + d(x_1-1,2)^+\mathbf{1}(x_1>2)\right] \\ &> \mu_1(h_1-h_2) - \xi_2 h_2 + \nu_2 d(x_1,0) \\ &\quad +(\nu_1+\mu_1)\left[\tilde{d}(1,2)^+\mathbf{1}(x_1=2) + d(x_1-1,2)^+\mathbf{1}(x_1>2)\right], \end{aligned} \qquad \text{(A.17)}$$

with the inequality following from $\mu_i > \xi_i$, $i = 1,2$.

The proof of part (ii) is by induction on x_2. We have $\bar{d}(1,1) > \tilde{d}(1,1) > \tilde{d}(1,2)$ with the second inequality following from Lemma 2.4(i). Therefore, assuming that $\tilde{d}(1,2) \geq 0$, we get from (A.4)

$$\tilde{d}(1,2) = \tilde{D} + (\mu_1+\xi_2+\mu_2)\tilde{d}(1,2), \qquad \text{(A.18)}$$

where

$$\tilde{D} = \xi_1(h_1-h_2) - \mu_2 h_2 + \nu_2\tilde{d}(1,1) + \nu_1\mu_2 g(0,3) \geq 0. \qquad \text{(A.19)}$$

Because $d(2,0) > d(1,1) > \tilde{d}(1,1)$, (A.17) and (A.19) yield $C(2) > \tilde{D}$ and we get $\hat{d}(2,1) > 0$ from (A.16) and Lemma 2.2. Substituting in (A.12) we get

$$d(2,1) = \mu_1(h_1-h_2) - \mu_2 h_2 + \nu_1 d(1,2) + \nu_2 d(2,0) + \mu_1\tilde{d}(1,2) + (\xi_1+\xi_2+\mu_2)d(2,1),$$

which combined with (A.18) and (A.19) gives

$$d(2,1) - \tilde{d}(1,2) > \frac{(\mu_1-\xi_1)(h_1-h_2) + \nu_1 d(1,2) + \xi_1\tilde{d}(1,2) + \nu_2\left[d(2,0)-\tilde{d}(1,1)\right]}{\nu_1+\nu_2+\mu_1},$$

which is positive because $d(1,2) > \tilde{d}(1,2) \geq 0$ and $d(2,0) > \tilde{d}(1,1)$, thus establishing the induction base. For $x_2 > 2$, assuming $\tilde{d}(1,x_2) \geq 0$ implies $\tilde{d}(1,x_2-1) \geq 0$ by Lemma 2.4(i), which by the induction hypothesis yields $d(2,x_2-2) \geq 0$. Taking into account all of the above we get from (A.13) and (A.4)

$$\begin{aligned} d(2,x_2-1) - \tilde{d}(1,x_2) \;&=\; (\mu_1 - \xi_1)(h_1 - h_2) + \nu_1 d(1,x_2) + \xi_1 \tilde{d}(1,x_2) - \nu_1 \mu_2 g(0,x_2+1) \\ &+\nu_2 \left[d(2,x_2-2) - \tilde{d}(1,x_2-1)\right] + (\xi_1+\xi_2)\left[d(2,x_2-1) - \tilde{d}(1,x_2)\right] \\ &+\mu_1 \left[d(2,x_2-1)^- - \tilde{d}(1,x_2)^-\right] + \mu_2 \left[d(2,x_2-1)^+ - \tilde{d}(1,x_2)^+\right]. \end{aligned}$$

Noting that $d(1,x_2) > \tilde{d}(1,x_2) \geq 0$ and applying the induction hypothesis to the term multiplying ν_2, we get from Lemma 2.2 that $d(2,x_2-1) - \tilde{d}(1,x_2) \geq 0$.

The proof of part (iii) is by nested induction on x_1 and x_2. For some $x_1 \geq 2$, assume that $d(x_1,2) \geq 0$. Then, $d(x_1,1) > 0$ by Lemma 2.5(i) and consequently $\hat{d}(x_1,1) > 0$ as well. Therefore, we get from (A.13)

$$d(x_1,2) = D(x_1) + (\xi_1+\xi_2+\mu_2)d(x_1,2), \tag{A.20}$$

where

$$\begin{aligned} D(x_1) \;&=\; \mu_1(h_1-h_2) - \mu_2 h_2 + \nu_1 d(x_1-1,3) + \nu_2 d(x_1,1) \\ &+\mu_1 \left[\tilde{d}(1,3)^+ \mathbf{1}(x_1=2) + d(x_1-1,3)^+ \mathbf{1}(x_1>2)\right] \geq 0. \end{aligned} \tag{A.21}$$

Assuming that the lemma holds for less than x_1 jobs in Station 1 (induction hypothesis with respect to x_1), we get $d(x_1,2) \geq d(x_1-1,3)$ if $x_1 > 2$. If $x_1 = 2$ we have $d(2,2) \geq d(1,3) > \tilde{d}(1,3)$, where the first inequality is due to part (i). Moreover, $d(x_1+1,0) > d(x_1,0) > d(x_1,1)$, so (A.17) and (A.21) yield $C(x_1+1) > D(x_1)$ and $\hat{d}(x_1+1,1) > 0$ follows from (A.16) and Lemma 2.2. Substituting in (A.12) we get

$$\begin{aligned} d(x_1+1,1) \;&=\; \mu_1(h_1-h_2) - \mu_2 h_2 + \nu_1 d(x_1,2) + \nu_2 d(x_1+1,0) \\ &+\mu_1 d(x_1,2) + (\xi_1+\xi_2+\mu_2)d(x_1+1,1). \end{aligned} \tag{A.22}$$

Using part (i) for $x_1 = 2$ and the induction hypothesis for $x_1 > 2$ we get from (A.20), (A.21) and (A.22) that $d(x_1,2) \leq d(x_1+1,1)$, which establishes the base for the induction with respect to x_2. To complete the induction we consider $x_1 \geq 2$, $x_2 \geq 3$, in which case (A.13) yields

$$\begin{aligned} d(x_1+1,x_2-1) - d(x_1,x_2) \;&=\; \nu_1\left[d(x_1,x_2) - d(x_1-1,x_2+1)\right] \\ +\nu_2\left[d(x_1+1,x_2-2) - d(x_1,x_2-1)\right] &+ \mu_1\left[d(2,x_2)^+ - \tilde{d}(1,x_2+1)^+\right]\mathbf{1}(x_1=2) \\ +\mu_1\left[d(x_1,x_2)^+ - d(x_1-1,x_2+1)^+\right]\mathbf{1}(x_1>2) &+ \mu_2\left[\hat{d}(x_1+1,1)^- - d(x_1,2)^-\right]\mathbf{1}(x_2=3) \\ +\mu_2\left[d(x_1+1,x_2-2)^- - d(x_1,x_2-1)^-\right]\mathbf{1}(x_2>3) &+ (\xi_1+\xi_2)\left[d(x_1+1,x_2-1) - d(x_1,x_2)\right] \\ +\mu_1\left[d(x_1+1,x_2-1)^- - d(x_1,x_2)^-\right] &+ \mu_2\left[d(x_1+1,x_2-1)^+ - d(x_1,x_2)^+\right]. \end{aligned} \tag{A.23}$$

Assume $d(x_1,x_2) \geq 0$. Then, reasoning as in the case $x_2 = 2$, we get $d(2,x_2) \geq d(1,x_2+1) > \tilde{d}(1,x_2+1)$ if $x_1 = 2$, and $d(x_1,x_2) \geq d(x_1-1,x_2+1)$ if $x_1 > 2$. Because $d(x_1,x_2) \geq 0$, by Lemma 2.5(i) we also have $d(x_1, x_2 - 1) \geq 0$ and applying the induction hypothesis with respect to x_2 we get $d(x_1, x_2 - 1) \leq d(x_1 + 1, x_2 - 2)$. Therefore, we get $d(x_1 + 1, x_2 - 1) \geq d(x_1, x_2)$ from (A.23) and Lemma 2.2.

Proof of Lemma 2.7

To prove part (i), we use the optimality equations to derive the following recursive equation for $f(1, x_2)$.

$$\begin{aligned} f(1,0) &= h_1 - h_2 + (\nu_2 + \xi_2)V(0,1) + (\nu_2 + \mu_1 + \mu_2 + \xi_2)f(1,0), && \text{(A.24)} \\ f(1,1) &= h_1 - h_2 + (\nu_2 + \mu_2)f(1,0) + (\mu_1 + \xi_2)f(1,1) \\ &\quad +(\mu_2 - \xi_2)\left[V(1,1) - V(1,0)\right], && \text{(A.25)} \\ f(1,x_2) &= h_1 - h_2 + (\nu_2 + \mu_2)f(1,x_2 - 1) + (\mu_1 + \xi_2)f(1,x_2), \quad x_2 \geq 2. && \text{(A.26)} \end{aligned}$$

From (A.24) and (A.25) we get

$$\begin{aligned} f(1,0) - f(1,1) &= (\nu_2 + \xi_2)V(0,1) + (\mu_1 + \mu_2 + \xi_2)\left[f(1,0) - f(1,1)\right] \\ &\quad -\mu_2\left[V(0,2) - V(0,1)\right] + \xi_2\left[V(1,1) - V(1,0)\right]. && \text{(A.27)} \end{aligned}$$

Using (2.3.5), (2.3.6) and $\nu_2 \geq \mu_2$ we get

$$(\nu_2 + \xi_2)V(0,1) - \mu_2\left[V(0,2) - V(0,1)\right] = h_2\left[1 - \frac{2\mu_2}{\nu_2 + \mu_2}\right] \geq 0,$$

which combined with (A.27) and Lemma 2.1 yields $f(1,0) > f(1,1)$. Then, $f(1,1) > f(1,2)$ follows from (A.25), (A.26), $\mu_2 > \xi_2$ and Lemma 2.1. Finally, $f(1,x_2) > f(1,x_2+1)$ for $x_2 \geq 2$ can be proved by induction on x_2 based on (A.26).

We now proceed to the proof of part (ii). Recall that condition $\mu_1 \geq \mu_2$ was not used in the proof of Lemma 2.5(i) for $x_1 \geq 2$, so the same arguments can be used here. Therefore, we only need to show that $d(1,x_2) > d(1,x_2+1)$. Equation (A.8) can be written as

$$d(1,0) = \mu_1(h_1 - h_2) + \mu_1\nu_2 V(0,1) + \mu_1\xi_2 V(1,0) + (\nu_2 + \mu_1 + \mu_2)d(1,0). \qquad \text{(A.28)}$$

Using the fact that $\bar{d}(1,1) = \xi_2 g(1,1)$ for $\xi_1 = 0$, we get after some straightforward algebra

$$\xi_2 d(1,1) + (\mu_1 - \mu_2)\bar{d}(1,1)^- = \mu_1\xi_2\left[V(1,0) - V(0,2)\right].$$

Substituting the last expression in (A.9) and combining with (A.28) we get

$$d(1,0) - d(1,1) = \frac{\mu_2 h_2 + \mu_1\nu_2 V(0,1) + \mu_1\xi_2 V(0,2) - \mu_2(\nu_1 - \mu_1)g(0,2)}{1 - \mu_1 - \mu_2} > 0. \qquad \text{(A.29)}$$

To prove $d(1,1) > d(1,2)$ we will assume $\mu_1 < \mu_2$ because for $\mu_1 \geq \mu_2$ it has been proved in Lemma 2.5(i). Setting $\bar{d}(1,1) = \xi_2 g(1,1)$ in (A.9) and (A.10) and $\tilde{d}(1,2) = \mu_2 g(1,2)$ in (A.10) we get after some algebraic manipulations

$$d(1,1) - d(1,2) = K + (\mu_1 + \xi_2)\left[d(1,1) - d(1,2)\right], \qquad \text{(A.30)}$$

where

$$
\begin{aligned}
K &= \nu_2\left[d(1,0)-d(1,1)\right]+\nu_1\mu_2\left[g(0,2)-g(0,3)\right]-\mu_1\mu_2 g(0,2) \\
&\quad +(\mu_1-\mu_2)\xi_2 g(1,1)+\mu_2(\mu_2-\xi_2)g(1,1). \qquad \text{(A.31)}
\end{aligned}
$$

Furthermore, it is straightforward to show that

$$
g(1,1)=V(1,0)-V(1,1)=\frac{-h_2+\nu_1 g(0,2)}{\nu_1+\nu_2+\xi_2}. \qquad \text{(A.32)}
$$

Then, noting that $\mu_1 < \mu_2$, $g(1,1) < 0$ and $g(0,2) > g(0,3)$, we substitute (A.29) and (A.32) into (A.31) to get

$$
\begin{aligned}
K &> \frac{\nu_2(\mu_1+\mu_2)-\mu_2(\mu_2-\xi_2)}{\nu_1+\nu_2+\xi_2}h_2 \\
&\quad -\frac{\nu_2\left[\mu_1\xi_2+\mu_2(\nu_1-\mu_1)\right]+\mu_1\mu_2(\nu_1+\nu_2+\xi_2)-\nu_1\mu_2(\mu_2-\xi_2)}{\nu_1+\nu_2+\xi_2}g(0,2), \qquad \text{(A.33)}
\end{aligned}
$$

where in (A.29) we have used $V(0,2) = V(0,1) - g(0,2)$ and $V(0,1) = h_2/(\nu_2 + \xi_2)$. The terms multiplying h_2 and $g(0,2)$ in (A.33) are positive because $\nu_2 \geq \mu_2$. Therefore, $K > 0$ and $d(1,1) > d(1,2)$ follows from (A.30). We prove $d(1,x_2) > d(1,x_2+1)$ for $x_2 \geq 2$ by induction on x_2 based on (A.10). It is straightforward to show that

$$
\mu_2 d(1,x_2)-\mu_1\mu_2 g(0,x_2+1)+(\mu_1-\mu_2)\mu_2 g(1,x_2)=\mu_1\mu_2 f(1,x_2-1),
$$

and (A.10) takes the form

$$
\begin{aligned}
d(1,x_2) &= \mu_1(h_1-h_2)-\mu_2 h_2+\nu_2 d(1,x_2-1)+(\mu_1+\xi_2)d(1,x_2) \\
&\quad +\mu_2\nu_1 g(0,x_2+1)+\mu_1\mu_2 f(1,x_2-1) \\
&\quad +\mu_2\left[\bar{d}(1,1)^{-}\mathbf{1}(x_2=2)+\tilde{d}(1,x_2-1)^{-}\mathbf{1}(x_2>2)\right], \quad x_2\geq 2. \qquad \text{(A.34)}
\end{aligned}
$$

Then, $d(1,x_2) - d(1,x_2+1) > 0$ is proved by induction on x_2, using the facts that $\bar{d}(1,1) > \tilde{d}(1,1)$ and $g(0,x_2)$, $f(1,x_2)$, and $\tilde{d}(1,x_2)$ are decreasing sequences.

As in the proof of part (ii), it suffices to show part (iii) for $x_1 = 1$. Let $L = \lim_{x_2\to\infty} d(1,x_2)$ and $F = \lim_{x_2\to\infty} f(1,x_2)$; F exists and is finite because $f(1,x_2)$ is a positive (Lemma 2.3) and decreasing sequence (part (i)). From (A.34) we have

$$
\begin{aligned}
d(1,x_2) &\leq \mu_1(h_1-h_2)-\mu_2 h_2+\nu_2 d(1,x_2-1)+(\mu_1+\xi_2)d(1,x_2) \\
&\quad +\mu_2\nu_1 g(0,x_2+1)+\mu_1\mu_2 f(1,x_2-1).
\end{aligned}
$$

Assuming L is finite we get a contradiction because by taking limits on both sides we get

$$
(\nu_1+\mu_2)L \leq \mu_1(h_1-h_2)-\mu_2 h_2+\mu_1\mu_2 F+\mu_2\nu_1\lim_{x_2\to\infty} g(0,x_2) = -\infty.
$$

Proof of Theorem 2.3

When there is no dedicated server assigned to Station 1, the optimality equations are given by (2.3.9) and (2.3.10) for all $x_1 \geq 1$ and $x_2 = 1$, $x_2 > 1$, respectively, with $\nu_1 = \xi_1 = 0$. Therefore, the decision function whose sign determines the optimal policy is $\hat{d}(x_1, 1)$ for one job downstream and $d(x_1, x_2)$ otherwise. Because $d(x_1, 1) < \hat{d}(x_1, 1)$, $x_1 \geq 1$, to prove that the optimal policy is characterized by a switching curve $t(x_1)$ it suffices to show that $d(x_1, x_2)$, $x_1 \geq 1$, is decreasing in x_2 and it becomes negative for x_2 large enough. We have

$$\begin{aligned} d(x_1, 0) &= \mu_1(h_1 - h_2) + (\nu_2 + \mu_2 + \xi_2) d(x_1, 0) \\ &\quad + \mu_1(\nu_2 + \xi_2) \left[V(x_1 - 1, 1) - V(x_1 - 1, 0) \right] \\ &\quad + \mu_1 \hat{d}(x_1 - 1, 1)^+ \mathbf{1}(x_1 > 1), \end{aligned} \tag{A.35}$$

$$\begin{aligned} d(x_1, 1) &= \mu_1(h_1 - h_2) - \mu_2 h_2 + \nu_2 d(x_1, 0) + (\mu_2 + \xi_2) d(x_1, 1) \\ &\quad + (\mu_1 - \mu_2) \hat{d}(x_1, 1)^- + \mu_1 d(x_1 - 1, 2)^+ \mathbf{1}(x_1 > 1), \end{aligned} \tag{A.36}$$

and for $x_2 \geq 2$

$$\begin{aligned} d(x_1, x_2) &= \mu_1(h_1 - h_2) - \mu_2 h_2 + \nu_2 d(x_1, x_2 - 1) + \xi_2 d(x_1, x_2) \\ &\quad + \mu_1 d(x_1, x_2)^- + \mu_2 d(x_1, x_2)^+ + \mu_1 d(x_1 - 1, x_2 + 1)^+ \mathbf{1}(x_1 > 1) \\ &\quad + \mu_2 \hat{d}(x_1, 1)^- \mathbf{1}(x_2 = 2) + \mu_2 d(x_1, x_2 - 1)^- \mathbf{1}(x_2 > 2). \end{aligned} \tag{A.37}$$

As part of the induction scheme used to prove the monotonicity of $d(x_1, x_2)$ we need to show $d(1, 0) > d(1, 1)$ and $d(x_1, 0) > d(x_1, 1)$, $x_1 > 1$, assuming that $d(x_1 - 1, x_2)$ is decreasing in x_2. When $\hat{d}(x_1, 1) \leq 0$ there is nothing to prove because $d(x_1, 1) < \hat{d}(x_1, 1)$ and $d(x_1, 0) > 0$ for all $x_1 \geq 1$. When $\hat{d}(1, 1) > 0$, $d(1, 0) > d(1, 1)$ follows directly from (A.35) and (A.36). For $x_1 > 1$ and $\hat{d}(x_1, 1) > 0$, taking into account Lemma 2.1, we get from (A.35) and (A.36)

$$\begin{aligned} d(x_1, 0) - d(x_1, 1) &> \mu_2 h_2 + (\mu_2 + \xi_2) \left[d(x_1, 0) - d(x_1, 1) \right] \\ &\quad + \mu_1 \left[\hat{d}(x_1 - 1, 1)^+ - d(x_1 - 1, 2)^+ \right]. \end{aligned}$$

The last term is positive because of the induction hypothesis, so $d(x_1, 0) - d(x_1, 1) > 0$. To prove that $d(1, 1) > d(1, 2)$ we consider $\hat{d}(1, 1) < 0$ and $\hat{d}(1, 1) \geq 0$ separately. When $\hat{d}(1, 1) < 0$, which implies $d(1, 1) < 0$ as well, we get from (A.36) and (A.37)

$$\begin{aligned} d(1, 1) - d(1, 2) &= \nu_2 d(1, 0) + (\mu_2 - \nu_2) d(1, 1) + \xi_2 \left[d(1, 1) - d(1, 2) \right] \\ &\quad + \mu_1 \left[\hat{d}(1, 1)^- - d(1, 2)^- \right] + \mu_2 \left[d(1, 1)^+ - d(1, 2)^+ \right] - 2\mu_2 \hat{d}(1, 1). \end{aligned}$$

Noting that $\hat{d}(1, 1) > d(1, 1)$, we can apply Lemma 2.2 to get $d(1, 1) > d(1, 2)$. Let now $\hat{d}(1, 1) \geq 0$. If $d(1, 1) \geq 0$ as well, by setting $\mu_2 d(1, 1) = \mu_2 d(1, 1)^+$ in (A.36) we get from (A.36) and (A.37)

$$\begin{aligned} d(1, 1) - d(1, 2) &= \nu_2 \left[d(1, 0) - d(1, 1) \right] + \xi_2 \left[d(1, 1) - d(1, 2) \right] \\ &\quad + \mu_2 \left[d(1, 1)^+ - d(1, 2)^+ \right] - \mu_1 d(1, 2)^-, \end{aligned}$$

and the result follows from the induction hypothesis and Lemma 2.2. If $d(1,1) < 0$, (A.36) and (A.37) give

$$\begin{aligned} d(1,1) - d(1,2) &= \nu_2 d(1,0) + (\mu_2 - \nu_2) d(1,1) + \xi_2 [d(1,1) - d(1,2)] \\ &\quad + \mu_2 \left[d(1,1)^+ - d(1,2)^+ \right] - \mu_1 d(1,2)^-, \end{aligned}$$

and the result follows from $\nu_2 \geq \mu_2$ and Lemma 2.2. The proof of $d(x_1, 1) > d(x_1, 2)$ for $x_1 > 1$ uses the same arguments because $d(x_1, 1) - d(x_1, 2)$ has the same form as $d(1,1) - d(1,2)$ with an additional term $\mu_1 [d(x_1 - 1, 2)^+ - d(x_1 - 1, 3)^+]$ which is positive by induction. For $x_2 \geq 2$, $d(x_1, x_2) > d(x_1, x_2 + 1)$ can be proved by a straightforward induction based on (A.37).

Because $d(x_1, x_2)$ is a decreasing sequence, $L(x_1) = \lim_{x_2 \to \infty} d(x_1, x_2)$ exists. We will use induction on x_1 to show that $L(x_1) < 0$ when $\mu_1(h_1 - h_2) < \mu_2 h_2$. Assuming that $L(1) \geq 0$ and taking limits in (A.37) we get $\mu_1 L(1) = \mu_1(h_1 - h_2) - \mu_2 h_2$, which is a contradiction. Assume now that $L(x_1 - 1) < 0$ and $L(x_1) \geq 0$. Then

$$L(x_1) = \frac{\mu_1(h_1 - h_2) - \mu_2 h_2}{\mu_1} + L(x_1 - 1),$$

again a contradiction.

To prove that the slope of $t(x_1)$ is at least -1 it suffices to show that for each $x_1 \geq 1$, $x_2 \geq 2$, $d(x_1, x_2) \geq 0$ implies $d(x_1, x_2) \leq d(x_1 + 1, x_2 - 1)$. For $x_1 \geq 2$, $\hat{d}(x_1, 1)$ is given by (A.16) and (A.17) with $\nu_1 = \xi_1 = 0$ and $\tilde{d}(1,2)$ replaced with $d(1,2)$. Then, the proof follows the same steps as that of part (iii) of Lemma 2.6.

To prove part (ii) we will use induction on x_1 to show that $L(x_1) \geq 0$ when $\mu_1(h_1 - h_2) \geq \mu_2 h_2$. Assuming that $d(1, x_2) < 0$ for x_2 sufficiently large we get from (A.37)

$$(\nu_2 + \mu_2) [d(1, x_2) - d(1, x_2 - 1)] = \mu_1(h_1 - h_2) - \mu_2 h_2 \geq 0,$$

which contradicts the monotonicity property of $d(1, x_2)$. Therefore, $L(1) \geq 0$. Assuming $L(x_1 - 1) \geq 0$ and repeating the argument we obtain

$$(\nu_2 + \mu_2) [d(x_1, x_2) - d(x_1, x_2 - 1)] = \mu_1(h_1 - h_2) - \mu_2 h_2 + \mu_1 d(x_1 - 1, x_2 + 1) \geq 0,$$

which is again a contradiction, thus completing the induction and the proof.

Proof of Theorem 2.4

When there is no dedicated server assigned to Station 2, the optimality equations are given by (2.3.8) and (2.3.10) for all $x_2 \geq 1$ and $x_1 = 1$, $x_1 > 1$, respectively, with $\nu_2 = \xi_2 = 0$. Therefore, the decision function whose sign determines the optimal policy is $\tilde{d}(1, x_2)$ for one job in the first station, and $d(x_1, x_2)$ otherwise.

To prove the existence of $t(x_1)$ it suffices to show that $\tilde{d}(1, x_2)$ and $d(x_1, x_2)$ are decreasing and their limit as $x_2 \to \infty$ is $-\infty$. For $\tilde{d}(1, x_2)$ we have

$$\begin{aligned} \tilde{d}(1,0) &= \xi_1(h_1 - h_2) + \xi_1 \mu_2 V(1,0) + \mu_1 \tilde{d}(1,0), && \text{(A.38)} \\ \tilde{d}(1,x_2) &= \xi_1(h_1 - h_2) - \mu_2 h_2 + \mu_1 \tilde{d}(1,x_2) + \nu_1 \mu_2 g(0, x_2 + 1) \\ &\quad + \xi_1 \tilde{d}(1,x_2)^- + \mu_2 \tilde{d}(1,x_2)^+ + \mu_2 \tilde{d}(1,x_2 - 1)^- \mathbf{1}(x_2 > 1), \quad x_2 \geq 1. && \text{(A.39)} \end{aligned}$$

When $\tilde{d}(1,1) < 0$, we get $\tilde{d}(1,0) > \tilde{d}(1,1)$ by subtracting (A.39) from (A.38) and taking into account that $g(0,2)$ is negative. For $\tilde{d}(1,1) \geq 0$, (A.39) becomes

$$\tilde{d}(1,1) = \xi_1(h_1 - h_2) - \mu_2 h_2 + (\mu_1 + \mu_2)\tilde{d}(1,1) + \nu_1 \mu_2 g(0, x_2 + 1). \tag{A.40}$$

Then, we get $\tilde{d}(1,0) > \tilde{d}(1,1)$ by writing (A.38) in the form $\tilde{d}(1,0) = \xi_1(h_1 - h_2) + \xi_1\mu_2 V(0,1) + (\mu_1 + \mu_2)\tilde{d}(1,0)$ and subtracting (A.40). For $x_2 \geq 1$, $\tilde{d}(1,x_2) - \tilde{d}(1,x_2+1) > 0$ follows from (A.39) by induction on x_2 and application of Lemma 2.2. To establish the limiting behavior of $\tilde{d}(1,x_2)$, note that for $x_2 > 2$ (A.39) is obtained from (A.4) by setting $\nu_2 = \xi_2 = 0$. Therefore, $\lim_{x_2 \to \infty} \tilde{d}(1,x_2) = -\infty$ by the arguments in the proof of Lemma 2.4(ii).

For $d(x_1,x_2)$ we get the following equations.

$$\begin{aligned} d(1,0) &= \mu_1(h_1 - h_2) + \mu_1\mu_2 V(0,1) + (\mu_1 + \mu_2)d(1,0), & \text{(A.41)} \\ d(1,x_2) &= \mu_1(h_1 - h_2) - \mu_2 h_2 + (\mu_1 + \mu_2)d(1,x_2) + \mu_2(\nu_1 + \xi_1 - \mu_1)g(0,x_2+1) \\ &\quad +(\mu_1 - \mu_2)\tilde{d}(1,x_2)^- + \mu_2\tilde{d}(1,x_2-1)^- \mathbf{1}(x_2 > 1), \quad x_2 \geq 1, & \text{(A.42)} \end{aligned}$$

and for $x_1 \geq 2$

$$\begin{aligned} d(x_1,0) &= \mu_1(h_1 - h_2) + \nu_1\mu_1 f(x_1-1,1) - \mu_1\mu_2 g(x_1-1,1) \\ &\quad +(\mu_2 + \xi_1)d(x_1,0) + \mu_1\left[\tilde{d}(1,1)^+\mathbf{1}(x_1 = 2)\right. \\ &\quad \left. +d(x_1-1,1)^+\mathbf{1}(x_1 > 2)\right], & \text{(A.43)} \\ d(x_1,x_2) &= \mu_1(h_1 - h_2) - \mu_2 h_2 + \nu_1 d(x_1-1,x_2+1) + \xi_1 d(x_1,x_2) \\ &\quad +\mu_1 d(x_1,x_2)^- + \mu_2 d(x_1,x_2)^+ + \mu_2 d(x_1,x_2-1)^-\mathbf{1}(x_2 > 1) \\ &\quad +\mu_1\left[\tilde{d}(1,x_2+1)^+\mathbf{1}(x_1 = 2)\right. \\ &\quad \left. +d(x_1-1,x_2+1)^+\mathbf{1}(x_1 > 2)\right], \quad x_2 \geq 1. & \text{(A.44)} \end{aligned}$$

Because $\mu_1 \geq \mu_2$ and $\tilde{d}(1,x_2)$ and $g(0,x_2)$ are decreasing sequences, $d(1,x_2) > d(1,x_2+1)$ can be easily proved by induction on x_2 based on (A.41) and (A.42). For $x_1 \geq 2$ we use induction on x_1. Because $d(x_1,0) > 0$, to show that $d(x_1,0) > d(x_1,1)$ we only need to consider $d(x_1,1) > 0$, in which case (A.44) gives

$$\begin{aligned} d(x_1,1) &= \mu_1(h_1 - h_2) - \mu_2 h_2 + \nu_1 d(x_1-1,2) + (\mu_2 + \xi_1)d(x_1,1) \\ &\quad +\mu_1\left[\tilde{d}(1,2)^+\mathbf{1}(x_1 = 2) + d(x_1-1,2)^+\mathbf{1}(x_1 > 2)\right]. & \text{(A.45)} \end{aligned}$$

The result follows in a straightforward manner from (A.43) and (A.45) because of $\tilde{d}(1,x_2)$ being decreasing ($x_1 = 2$) or the induction hypothesis ($x_1 > 2$). For $x_2 \geq 1$, $d(x_1,x_2) > d(x_1,x_2+1)$ is proved by induction on x_2 based on (A.44) with Lemma 2.2 used as well. Finally, because for $x_2 > 2$ (A.42) and (A.44) are special cases of (A.10) and (A.13) for $\nu_2 = \xi_2 = 0$, $\lim_{x_2 \to \infty} d(x_1,x_2) = -\infty$ is proved exactly as in Lemma 2.5(ii).

The fact that the slope of $t(x_1)$ is at least -1 is a consequence of the properties cited in the statement of Lemma 2.6. Their proof is very similar and even simpler because $\hat{d}(x_1,1)$ is not involved, so we omit the details.

Having shown that $\tilde{d}(1, x_2)$ and $d(x_1, x_2)$ are decreasing sequences, to prove part (ii) it suffices to show that $\tilde{d}(1, 1) < 0$ and $d(x_1, 1) < 0$ when $\mu_1(h_1 - h_2) \le \mu_2 h_2$. Assuming $\tilde{d}(1, 1) \ge 0$, (A.40) leads to a contradiction. Because $\mu_1 \ge \mu_2$ and $\tilde{d}(1, 1) < 0$, $d(1, 1) < 0$ follows from (A.42). For $x_1 \ge 2$ we prove $d(x_1, 1) < 0$ by induction on x_1. Assuming that $d(x_1 - 1, x_2) < 0$ for $x_2 \ge 1$ (induction hypothesis) and $d(x_1, 1) \ge 0$ we get a contradiction from applying (A.44), and the proof is complete.

Proof of Theorem 2.5

The first part is a direct consequence of Proposition 2.2 and Lemma 2.3. For the second part we only need to show that $d(1, x_2)$ is decreasing and its limit is equal to $-\infty$; the proof of these properties for $x_1 \ge 2$ is identical to the corresponding part of the proof of Theorem 2.4 because it did not use condition $\mu_1 \ge \mu_2$. For $x_1 = 1$ we have

$$\begin{aligned} d(1,0) &= \mu_1(h_1 - h_2) + \mu_1\mu_2 V(0,1) + (\mu_1 + \mu_2) d(1,0), \\ d(1,x_2) &= \mu_1(h_1 - h_2) - \mu_2 h_2 + \mu_1 d(1,x_2) + \mu_2 d(1,x_2 - 1) \\ &\quad + \mu_2\nu_1 g(0, x_2 + 1), \quad x_2 \ge 1, \end{aligned} \tag{A.46}$$

where (A.46) was obtained by setting $\xi_1 = 0$ and $\tilde{d}(1, x_2)^- = [\mu_2 g(1, x_2)]^- = \mu_2 g(1, x_2)$ in (A.42). The monotonicity of $d(1, x_2)$ is proved by a straightforward induction on x_2, while its limiting property is a consequence of the analogous limiting behavior of $g(0, x_2)$.

Proof of Theorem 2.8

When $f(x_1, x_2) < 0$, Propositions 2.1 and 2.2 imply that the dedicated server at Station 1 should be idled and the flexible server should be assigned to Station 2. Therefore, to prove the first part of the theorem we only need to show that decision functions $\tilde{d}(1, x_2)$ and $d(x_1, x_2)$ are negative for $x_1 \ge 2$, $x_2 \ge 1$ such that $f(x_1, x_2) \ge 0$.

Because $f(1, x_2)$ is a decreasing sequence (see proof of Theorem 2.6), there exists x_2^* such that $f(1, x_2) \ge 0$ for $x_2 \le x_2^*$. If $x_2^* = 0$, there is nothing to prove. Otherwise, the expression we have derived for $\tilde{d}(1, x_2)$ when $h_1 \ge h_2$ (Equation A.39) is valid for $x_2 \le x_2^*$, that is,

$$\begin{aligned} \tilde{d}(1,x_2) &= \xi_1(h_1 - h_2) - \mu_2 h_2 + \mu_1 \tilde{d}(1,x_2) + \nu_1\mu_2 g(0, x_2+1) \\ &\quad + \xi_1 \tilde{d}(1,x_2)^- + \mu_2 \tilde{d}(1,x_2)^+ + \mu_2 \tilde{d}(1,x_2-1)\mathbf{1}(x_2 > 1), \quad 1 \le x_2 \le x_2^*, \end{aligned}$$

and $\tilde{d}(1, x_2) < 0$ can be proved by a straightforward induction on x_2. Therefore, the optimal allocation in state $(1, x_2)$, $x_2 \le x_2^*$, is (ν_1, μ_2), and optimality equations (2.3.1) and (2.3.2) give

$$d(1,1) = \mu_1(h_1 - h_2) - \mu_2 h_2 + (\mu_1 + \xi_1) d(1,1) + \nu_1\mu_2 g(0,2) + \mu_2(\mu_1 - \xi_1) f(1,0), \tag{A.47}$$

and for $x_2 \ge 2$,

$$d(1,x_2) = \mu_1(h_1 - h_2) - \mu_2 h_2 + (\mu_1 + \xi_1) d(1,x_2) + \nu_1\mu_2 g(0, x_2+1) + \mu_2 d(1, x_2 - 1). \tag{A.48}$$

From (2.3.3) and (2.3.5) we have $f(1,0) = h_1/(\nu_1+\xi_1)$. Because $h_1 < h_2$ and $\nu_1+\xi_1 > \mu_1$ it is easily seen that $(\mu_1-\xi_1)f(1,0)-h_2 < 0$, and $d(1,1) < 0$ follows from (A.47). For $x_2 \geq 2$, $d(1,x_2) < 0$ follows directly by applying induction in (A.48).

Next we show $d(x_1,x_2) < 0$ for $x_1 \geq 2$ by nested induction on x_1 and x_2. Assume that $d(x_1-1,x_2) < 0$ for $x_2 \geq 1$ and $d(x_1,x_2-1) < 0$ if $x_2 > 1$ (induction hypothesis). Note that for states (y_1,y_2) with $d(y_1,y_2) < 0$ the optimal allocation is either (ν_1,μ_2) or $(0,\mu_2)$, resulting in the following optimality equation.

$$\begin{aligned} V(y_1,y_2) &= h_1y_1 + h_2y_2 + \mu_2 V(y_1,y_2-1) \\ &\quad + \min\{\nu_1 V(y_1-1,y_2+1) + (\mu_1+\xi_1)V(y_1,y_2), (\nu_1+\mu_1+\xi_1)V(y_1,y_2)\}. \end{aligned} \quad \text{(A.49)}$$

Assuming $f(x_1,x_2) \geq 0$, we use (2.3.10) for $V(x_1,x_2)$ and (A.49) for $V(x_1-1,x_2+1)$ and $V(x_1,x_2-1)$ to get an expression for $d(x_1,x_2)$. Noting that the difference of the two terms in braces in (A.49) is equal to $\nu_1 f(y_1,y_2)$, we get for $x_2 \geq 1$

$$\begin{aligned} d(x_1,x_2) &= \mu_1(h_1-h_2) - \mu_2 h_2 + \nu_1\mu_1 f(x_1-1,x_2+1)^+ + \nu_1\mu_2 g(x_1-1,x_2+1) \\ &\quad +\xi_1 d(x_1,x_2) + \mu_1 d(x_1,x_2)^- + \mu_2 d(x_1,x_2)^+ \\ &\quad + \left[\mu_2 d(x_1,x_2-1) + \nu_1\mu_2 f(x_1,x_2-1)^-\right] \mathbf{1}(x_2 > 1). \end{aligned} \quad \text{(A.50)}$$

Because $\mu_1 f(x_1-1,x_2+1)^+ + \mu_2 g(x_1-1,x_2+1) = \max\{d(x_1-1,x_2+1), \mu_2 g(x_1-1,x_2+1)\} < 0$, we obtain $d(x_1,x_2) < 0$ by applying the induction hypothesis and Lemma 2.2 in (A.50).

For part (ii) it suffices to show that $f(x_1,x_2)$ is increasing in x_1 and decreasing in x_2. Taking into account that the optimal policy assigns the flexible server to the downstream station, we use (A.49) to derive the following recursive equation for $f(x_1,x_2)$, $x_1 \geq 1$, $x_2 \geq 1$.

$$\begin{aligned} f(x_1,x_2) &= h_1 - h_2 + (\mu_1+\xi_1)f(x_1,x_2) + \mu_2 f(x_1,x_2-1) \\ &\quad +\nu_1 f(x_1,x_2)^- + \nu_1 f(x_1-1,x_2+1)^+ \mathbf{1}(x_1 > 1). \end{aligned} \quad \text{(A.51)}$$

The proof of the first monotonicity property, $f(x_1+1,x_2) > f(x_1,x_2)$, $x_1 \geq 1$, is by a straightforward induction on x_1, x_2 based on (A.51), with the induction base for each $x_1 \geq 1$ established by the fact that $f(x_1+1,0) > f(x_1,0)$. The second monotonicity property has already been proved for $x_1 = 1$ (see proof of Theorem 2.6). For $x_1 \geq 2$ we can prove that $f(x_1,x_2)$ is decreasing in x_2 by similar induction arguments provided that we can also show that $f(x_1,1) < f(x_1,0)$ to establish the induction base for each x_1. For that purpose we use a sample path argument. Let $P1$ and $P2$ be the processes that start at states $(x_1,1)$ and $(x_1-1,2)$, respectively, and assume that the optimal policy, say π, is applied to $P2$. As for $P1$, we apply a policy $\bar{\pi}$ that imitates π until the first time that Station 2 is empty under $P1$ and has one job under $P2$, and is optimal afterwards. Let τ be that time and y_1 be the number of jobs in Station 1 under $P1$ at time τ. The two policies have a holding cost rate difference of $h_1 - h_2$ until time τ and are optimal afterwards. Therefore, because $\bar{\pi}$ is not necessarily optimal we have

$$V(x_1,1) - V(x_1-1,2) \leq (h_1-h_2)E(\tau) + E\left[V(y_1,0) - V(y_1-1,1)\right] < E\left[f(y_1,0)\right], \quad \text{(A.52)}$$

because $h_1 < h_2$. Along every sample path we have $y_1 \leq x_1$, so $f(y_1,0) \leq f(x_1,0)$, which combined with (A.52) yields $f(x_1,1) < f(x_1,0)$.

Proof of Lemma 2.8

First, we show the monotonicity of the value function with respect to x_2. The proof is by induction on n. Let (ρ_1^*, ρ_2^*) be the optimal allocations in state (x_1, x_2). Assuming a suboptimal allocation $(\rho_1^*, 0)$ for state $(x_1, x_2 - 1)$ we get from (2.3.15) and (2.3.16)

$$\begin{aligned} V_{n,\beta}(x_1,x_2) - V_{n,\beta}(x_1,x_2-1) \quad &\geq \quad h_2 + \beta \left[T_{\rho_1^*,\rho_2^*} V_{n-1,\beta}(x_1,x_2) - T_{\rho_1^*,0} V_{n-1,\beta}(x_1,x_2-1)\right] \\ &= h_2 + \beta \{\lambda \left[V_{n-1,\beta}(x_1+1,x_2) - V_{n-1,\beta}(x_1+1,x_2-1)\right] \\ &+ \rho_1^* \left[V_{n-1,\beta}(x_1-1,x_2+1) - V_{n-1,\beta}(x_1-1,x_2)\right] \\ &+ (1-\lambda-\rho_1^*-\rho_2^*) \left[V_{n-1,\beta}(x_1,x_2) - V_{n-1,\beta}(x_1,x_2-1)\right]\} \geq 0, \end{aligned}$$

by the induction hypothesis. Next, we show the monotonicity property with respect to x_1 by a similar argument. We have

$$\begin{aligned} V_{n,\beta}(x_1,x_2) - V_{n,\beta}(x_1-1,x_2) \quad &\geq \quad h_1 + \beta \left[T_{\rho_1^*,\rho_2^*} V_{n-1,\beta}(x_1,x_2) - T_{0,\rho_2^*} V_{n-1,\beta}(x_1-1,x_2)\right] \\ &= h_1 + \beta \{\lambda \left[V_{n-1,\beta}(x_1+1,x_2) - V_{n-1,\beta}(x_1,x_2)\right] \\ &+ \rho_1^* \left[V_{n-1,\beta}(x_1-1,x_2+1) - V_{n-1,\beta}(x_1-1,x_2)\right] \\ &+ \rho_2^* \left[V_{n-1,\beta}(x_1,x_2-1) - V_{n-1,\beta}(x_1-1,x_2-1)\right] \\ &+ (1-\lambda-\rho_1^*-\rho_2^*) \left[V_{n-1,\beta}(x_1,x_2) - V_{n-1,\beta}(x_1-1,x_2)\right]\} \geq 0, \end{aligned}$$

because of the monotonicity with respect to x_2 and the induction hypothesis.

Proof of Lemma 2.9

The proof is by induction on n. Let (ρ_1^*, ρ_2^*) be the optimal allocations in state (x_1, x_2). Assuming a suboptimal allocation $(0, \rho_2^*)$ for state $(x_1 - 1, x_2 + 1)$ we get from (2.3.15) and (2.3.16)

$$\begin{aligned} f_{n,\beta}(x_1,x_2) \quad &\geq \quad h_1 - h_2 + \beta \left[\lambda f_{n-1,\beta}(x_1+1,x_2) + \rho_2^* f_{n-1,\beta}(x_1,x_2-1)\right. \\ &\left. + (1-\lambda-\rho_1^*-\rho_2^*) f_{n-1,\beta}(x_1,x_2)\right] \geq 0, \end{aligned}$$

by the induction hypothesis.

Proof of Theorem 2.9

For part (i) we only need to consider $x_2 \geq 2$ because $Y_1 \geq 1$. The proof is by sample path arguments, similar to those used for the proof of Theorem 4.2 in [64]. We denote by $P1$ and $P2$ the processes that start at states (x_1, x_2) and $(x_1 - 1, x_2 + 1)$, respectively, and assume that the optimal policy, say π, is applied to $P2$. As for $P1$, we add in Station 2 a fictitious job that incurs no cost and we apply a policy $\bar{\pi}$ that mimics π. Under $\bar{\pi}$ the fictitious job has the lowest priority[1] so that it has no effect on the cost incurred by real

[1]The fictitious job is assigned a server when there is no other job that has not been assigned a server and is preempted by an arriving job from Station 1.

jobs. Policies $\bar{\pi}$ and π have a cost difference $h_1 - h_2$ per period as long as the fictitious job is present, and h_1 after its departure. For a realization ω, let $T(\omega)$ be the time the fictitious job departs and $C_n^{\gamma}(\omega)$ be the cost incurred by a policy γ over a horizon of length n. Then, letting $S(\omega) = \min\{T(\omega), n\}$, we have

$$\begin{aligned} C_n^{\bar{\pi}}(\omega) - C_n^{\pi}(\omega) &= (h_1 - h_2) \sum_{k=0}^{S(\omega)-1} \beta^k + h_1 \sum_{k=S(\omega)}^{n-1} \beta^k \\ &= (h_1 - h_2)\frac{1-\beta^{S(\omega)}}{1-\beta} + h_1 \frac{\beta^{S(\omega)} - \beta^n}{1-\beta} = (h_1 - h_2)\frac{1-\beta^n}{1-\beta} + h_2 \frac{\beta^{S(\omega)} - \beta^n}{1-\beta}. \quad \text{(A.53)} \end{aligned}$$

Noting that the fictitious job leaves the system after at least x_2 service completions at Station 2 and at most one such event can happen at each period, we have $T(\omega) \geq x_2$. Therefore, either $S(\omega) = n$ or $x_2 \leq S(\omega) < n$. In both cases, taking also into account that $h_1 \geq h_2$, (A.53) yields

$$C_n^{\bar{\pi}}(\omega) - C_n^{\pi}(\omega) < \frac{h_1 - h_2 + h_2 \beta^{x_2}}{1-\beta}, \quad \text{(A.54)}$$

from which we get

$$f_{n,\beta}(x_1, x_2) = V_{n,\beta}(x_1, x_2) - V_{n,\beta}(x_1 - 1, x_2 + 1) < \frac{h_1 - h_2 + h_2 \beta^{x_2}}{1-\beta}, \quad \text{(A.55)}$$

because $\bar{\pi}$ is not necessarily optimal and (A.54) holds for expected values as well.

Applying the same sample path argument to the processes that start at states $(x_1, x_2 - 1)$ and (x_1, x_2), and defining policies π, $\bar{\pi}$ and time $S(\omega)$ accordingly, we obtain

$$C_n^{\bar{\pi}}(\omega) - C_n^{\pi}(\omega) = -h_2 \sum_{k=0}^{S(\omega)-1} \beta^k = -h_2 \frac{1-\beta^{S(\omega)}}{1-\beta}.$$

For $n \geq x_2$ we have $S(\omega) \geq x_2 - 1$, which leads to

$$C_n^{\bar{\pi}}(\omega) - C_n^{\pi}(\omega) \leq \frac{-h_2 + h_2 \beta^{x_2 - 1}}{1-\beta},$$

from which we get

$$g_{n,\beta}(x_1, x_2) = V_{n,\beta}(x_1, x_2 - 1) - V_{n,\beta}(x_1, x_2) \leq \frac{-h_2 + h_2 \beta^{x_2 - 1}}{1-\beta}. \quad \text{(A.56)}$$

Combining (A.55) and (A.56) we get

$$d_{n,\beta}(x_1, x_2) < \frac{\mu_1(h_1 - h_2) - \mu_2 h_2}{1-\beta} + \frac{\mu_1 h_2 \beta^{x_2} + \mu_2 h_2 \beta^{x_2 - 1}}{1-\beta}, \quad n \geq x_2. \quad \text{(A.57)}$$

Recalling that $Y_1 = \min\left\{x | \beta^x < \frac{(\mu_1+\mu_2)h_2 - \mu_1 h_1}{(\mu_1+\mu_2)h_2}\right\}$, it is clear from (A.57) that $d_{n,\beta}(x_1, x_2) < 0$ for $x_2 \geq Y_1 + 1$ and $n \geq x_2$. Because $\tilde{d}_{n,\beta}(1, x_2) \leq d_{n,\beta}(1, x_2)$, taking the limits as $n \to \infty$ we get $\tilde{d}_{\beta}(1, x_2) < 0$ and $d_{\beta}(x_1, x_2) < 0$, $x_1 \geq 2$, which proves part (i).

To prove part (ii) we need an expression for decision function $\hat{d}_{n,\beta}(x_1, 1)$. The arguments used in part (i) to get (A.55) are valid for $x_2 = 1$ as well. However, in the arguments used to get (A.56) we have $S(\omega) \geq 1$ for $x_2 = 1$. Therefore, $g_{n,\beta}(x_1, 1) \leq (-h_2 + h_2\beta)/(1-\beta)$, which combined with (A.55) yields

$$\hat{d}_{n,\beta}(x_1, 1) < \frac{\mu_1(h_1 - h_2) - \xi_2 h_2 + \beta(\mu_1 + \xi_2)h_2}{1-\beta},$$

which is negative for $\beta < \frac{(\mu_1+\xi_2)h_2 - \mu_1 h_1}{(\mu_1+\xi_2)h_2}$, implying $\bar{d}_{n,\beta}(1,1) < 0$ as well, so the optimal policy assigns the flexible server to Station 2 whenever there is one job there. For the aforementioned range for β we also have $Y_1 = 1$, so part (i) applies and the proof of part (ii) is complete.

For part (iii) we need to prove that $f_{n,\beta}(x_1, x_2) < 0$ for $x_2 \geq Y_2 - \mathbf{1}(Y_2 = 1)$. For that it suffices to show that $C_n^{\bar{\pi}}(\omega) - C_n^{\pi}(\omega)$ is negative along every sample path. We consider all $x_2 \geq 0$, so $T(\omega) \geq \max\{1, x_2\}$. For ω such that $S(\omega) = n$, (A.53) implies $C_n^{\bar{\pi}}(\omega) - C_n^{\pi}(\omega) < 0$ because $h_1 < h_2$. For ω such that $\max\{1, x_2\} \leq S(\omega) < n$ we get from (A.53)

$$C_n^{\bar{\pi}}(\omega) - C_n^{\pi}(\omega) < \frac{h_1 - h_2 + h_2\beta^{\max\{1,x_2\}}}{1-\beta}. \tag{A.58}$$

It is clear that $x_2 \geq Y_2 - \mathbf{1}(Y_2 = 1)$ implies $\max\{1, x_2\} \geq Y_2$. Therefore, $h_2\beta^{\max\{1,x_2\}} \leq h_2\beta^{Y_2} < h_2 - h_1$, and $C_n^{\bar{\pi}}(\omega) - C_n^{\pi}(\omega) < 0$ follows from (A.58).

Proof of Theorem 2.10

Because non-idling policies are optimal for $h_1 \geq h_2$, when there is no dedicated server in Station 1 optimality equation (2.3.15) takes the following form.

$$\begin{aligned} V_{n,\beta}(0,1) &= h_2 + \beta\left[\lambda V_{n-1,\beta}(1,1) + (\nu_2+\xi_2)V_{n-1,\beta}(0,0)\right. \\ &\quad \left.+(\mu_1+\mu_2)V_{n-1,\beta}(0,1)\right], \qquad (A.59) \\ V_{n,\beta}(0,x_2) &= h_2x_2 + \beta\left[\lambda V_{n-1,\beta}(1,x_2) + (\nu_2+\mu_2)V_{n-1,\beta}(0,x_2-1)\right. \\ &\quad \left.+(\mu_1+\xi_2)V_{n-1,\beta}(0,x_2)\right], \quad x_2 \geq 2, \qquad (A.60) \end{aligned}$$

and for $x_1 \geq 1$

$$\begin{aligned} V_{n,\beta}(x_1,0) &= h_1x_1 + \beta\left[\lambda V_{n-1,\beta}(x_1+1,0) + \mu_1 V_{n-1,\beta}(x_1-1,1)\right. \\ &\quad \left.+(\nu_2+\mu_2+\xi_2)V_{n-1,\beta}(x_1,0)\right], \qquad (A.61) \\ V_{n,\beta}(x_1,1) &= h_1x_1 + h_2 + \beta\left[\lambda V_{n-1,\beta}(x_1+1,1) + \nu_2 V_{n-1,\beta}(x_1,0) + \mu_2 V_{n-1,\beta}(x_1,1)\right. \\ &\quad \left.+\min\left\{\mu_1 V_{n-1,\beta}(x_1-1,2) + \xi_2 V_{n-1,\beta}(x_1,1), \xi_2 V_{n-1,\beta}(x_1,0) + \mu_1 V_{n-1,\beta}(x_1,1)\right\}\right], \qquad (A.62) \\ V_{n,\beta}(x_1,x_2) &= h_1x_1 + h_2x_2 + \beta\left[\lambda V_{n-1,\beta}(x_1+1,x_2) + \nu_2 V_{n-1,\beta}(x_1,x_2-1) + \xi_2 V_{n-1,\beta}(x_1,x_2)\right. \\ &\quad +\min\left\{\mu_1 V_{n-1,\beta}(x_1-1,x_2+1) + \mu_2 V_{n-1,\beta}(x_1,x_2),\right. \\ &\quad \left.\left.\mu_2 V_{n-1,\beta}(x_1,x_2-1) + \mu_1 V_{n-1,\beta}(x_1,x_2)\right\}\right], \quad x_2 \geq 2. \qquad (A.63) \end{aligned}$$

Replacing $V_{n,\beta}$ and $V_{n-1,\beta}$ with V_β in (A.59)-(A.63) we get the infinite horizon optimality equations from which it follows that the decision function whose sign determines the optimal allocation of the flexible server is $\hat{d}_\beta(x_1,1)$ for one job downstream and $d_\beta(x_1,x_2)$ otherwise. We have already shown that $d_\beta(x_1,x_2)$ becomes negative for x_2 sufficiently large (Theorem 2.9(i)). Therefore, taking into account that $d_\beta(x_1,1) \le \hat{d}_\beta(x_1,1)$ and Proposition 2.3, for part (i) it would suffice to show that $d_{n,\beta}(x_1,x_2)$ is non-increasing in x_2. From (A.59)-(A.63) we get

$$\begin{aligned} d_{n,\beta}(x_1,0) \;=\;& \mu_1(h_1-h_2)+\beta\left[\lambda d_{n-1,\beta}(x_1+1,0)+(\nu_2+\mu_2+\xi_2)d_{n-1,\beta}(x_1,0)\right.\\ &+\mu_1(\nu_2+\xi_2)\left[V_{n-1,\beta}(x_1-1,1)-V_{n-1,\beta}(x_1-1,0)\right]\\ &\left.+\mu_1\hat{d}_{n-1,\beta}(x_1-1,1)^+\mathbf{1}(x_1>1)\right], \end{aligned} \tag{A.64}$$

$$\begin{aligned} d_{n,\beta}(x_1,1) \;=\;& \mu_1(h_1-h_2)-\mu_2h_2+\beta\left[\lambda d_{n-1,\beta}(x_1+1,1)\right.\\ &+\nu_2 d_{n-1,\beta}(x_1,0)+(\mu_2+\xi_2)d_{n-1,\beta}(x_1,1)\\ &\left.+(\mu_1-\mu_2)\hat{d}_{n-1,\beta}(x_1,1)^-+\mu_1 d_{n-1,\beta}(x_1-1,2)^+\mathbf{1}(x_1>1)\right], \end{aligned} \tag{A.65}$$

and for $x_2 \ge 2$

$$\begin{aligned} d_{n,\beta}(x_1,x_2) \;=\;& \mu_1(h_1-h_2)-\mu_2h_2+\beta\left[\lambda d_{n-1,\beta}(x_1+1,x_2)\right.\\ &+\nu_2 d_{n-1,\beta}(x_1,x_2-1)+\xi_2 d_{n-1,\beta}(x_1,x_2)\\ &+\mu_1 d_{n-1,\beta}(x_1,x_2)^-+\mu_2 d_{n-1,\beta}(x_1,x_2)^++\mu_1 d_{n-1,\beta}(x_1-1,x_2+1)^+\mathbf{1}(x_1>1)\\ &\left.+\mu_2\hat{d}_{n-1,\beta}(x_1,1)^-\mathbf{1}(x_2=2)+\mu_2 d_{n-1,\beta}(x_1,x_2-1)^-\mathbf{1}(x_2>2)\right]. \end{aligned} \tag{A.66}$$

We use induction on n. We start by proving that $d_{n,\beta}(x_1,0) \ge d_{n,\beta}(x_1,1)$. When $\hat{d}_{n-1,\beta}(x_1,1) \ge 0$, $d_{n,\beta}(x_1,0) \ge d_{n,\beta}(x_1,1)$ follows directly from (A.64) and (A.65) by the monotonicity of the value function (Lemma 2.8) and the induction hypothesis. When $\hat{d}_{n-1,\beta}(x_1,1) < 0$, we get from (A.64) and (A.65)

$$\begin{aligned} d_{n,\beta}(x_1,0)-d_{n,\beta}(x_1,1) \;\ge\;& \mu_2h_2+\beta\left\{\lambda\left[d_{n-1,\beta}(x_1+1,0)-d_{n-1,\beta}(x_1+1,1)\right]\right.\\ &+\mu_2 d_{n-1,\beta}(x_1,0)+\xi_2\left[d_{n-1,\beta}(x_1,0)-d_{n-1}(x_1,1)\right]\\ &+\mu_2\left[\hat{d}_{n-1,\beta}(x_1,1)-d_{n-1,\beta}(x_1,1)\right]-\mu_1\hat{d}_{n-1,\beta,}(x_1,1)\\ &\left.+\mu_1\left[\hat{d}_{n-1,\beta}(x_1-1,1)^+-d_{n-1,\beta}(x_1-1,2)^+\right]\mathbf{1}(x_1>1)\right\}, \end{aligned}$$

which is nonnegative because of the induction hypothesis, $d_{n-1,\beta}(x_1,0) \ge 0$, and $d_{n-1,\beta}(x_1,1) \le \hat{d}_{n-1,\beta}(x_1,1)$. To prove that $d_{n,\beta}(x_1,1) \ge d_{n,\beta}(x_1,2)$ we again consider $\hat{d}_{n-1,\beta}(x_1,1) < 0$ and $\hat{d}_{n-1,\beta}(x_1,1) \ge 0$ separately. When $\hat{d}_{n-1,\beta}(x_1,1) < 0$, we also have $d_{n-1,\beta}(x_1,1) < 0$ and $d_{n-1,\beta}(x_1,2) < 0$ by the induction hypothesis. Then, we get from (A.65) and (A.66)

$$\begin{aligned} d_{n,\beta}(x_1,1)-d_{n,\beta}(x_1,2) \;=\;& \beta\left\{\lambda\left[d_{n-1,\beta}(x_1+1,1)-d_{n-1,\beta}(x_1+1,2)\right]+\nu_2 d_{n-1,\beta}(x_1,0)\right.\\ &+(\mu_2-\nu_2)d_{n-1,\beta}(x_1,1)+\xi_2\left[d_{n-1,\beta}(x_1,1)-d_{n-1,\beta}(x_1,2)\right]\\ &+\mu_1\left[\hat{d}_{n-1,\beta}(x_1,1)^--d_{n-1,\beta}(x_1,2)^-\right]-2\mu_2\hat{d}_{n-1,\beta}(x_1,1)\\ &\left.+\mu_1\left[d_{n-1,\beta}(x_1-1,2)^+-d_{n-1,\beta}(x_1-1,3)^+\right]\mathbf{1}(x_1>1)\right\} \ge 0, \end{aligned}$$

because of the induction hypothesis, $\nu_2 \geq \mu_2$, and $d_{n-1,\beta}(x_1,1) \leq \hat{d}_{n-1,\beta}(x_1,1) < 0$. Let now $\hat{d}_{n-1,\beta}(x_1,1) \geq 0$. If $d_{n-1,\beta}(x_1,1) \geq 0$ as well, we get from (A.65) and (A.66)

$$\begin{aligned} d_{n,\beta}(x_1,1) - d_{n,\beta}(x_1,2) &= \beta \{ \lambda [d_{n-1,\beta}(x_1+1,1) - d_{n-1,\beta}(x_1+1,2)] \\ &\quad + \nu_2 [d_{n-1,\beta}(x_1,0) - d_{n-1,\beta}(x_1,1)] + \xi_2 [d_{n-1,\beta}(x_1,1) - d_{n-1,\beta}(x_1,2)] \\ &\quad + \mu_2 \left[d_{n-1,\beta}(x_1,1)^+ - d_{n-1,\beta}(x_1,2)^+ \right] - \mu_1 d_{n-1,\beta}(x_1,2)^- \\ &\quad + \mu_1 \left[d_{n-1,\beta}(x_1-1,2)^+ - d_{n-1,\beta}(x_1-1,3)^+ \right] \mathbf{1}(x_1 > 1) \right\}, \end{aligned}$$

and the result follows from the induction hypothesis. If $d_{n-1,\beta}(x_1,1) < 0$, in which case the induction hypothesis implies that $d_{n-1,\beta}(x_1,2) < 0$ as well, (A.65) and (A.66) give

$$\begin{aligned} d_{n,\beta}(x_1,1) - d_{n,\beta}(x_1,2) &= \beta \{ \lambda [d_{n-1,\beta}(x_1+1,1) - d_{n-1,\beta}(x_1+1,2)] + \nu_2 d_{n-1,\beta}(x_1,0) \\ &\quad + (\mu_2 - \nu_2) d_{n-1,\beta}(x_1,1) + \xi_2 [d_{n-1,\beta}(x_1,1) - d_{n-1,\beta}(x_1,2)] - \mu_1 d_{n-1,\beta}(x_1,2)^- \\ &\quad + \mu_1 \left[d_{n-1,\beta}(x_1-1,2)^+ - d_{n-1,\beta}(x_1-1,3)^+ \right] \mathbf{1}(x_1 > 1) \right\}, \end{aligned}$$

and the result follows from the induction hypothesis, $\nu_2 \geq \mu_2$, and $d_{n-1,\beta}(x_1,1) < 0$. For $x_2 \geq 2$, $d_{n,\beta}(x_1,x_2) \geq d_{n,\beta}(x_1,x_2+1)$ follows by applying the induction hypothesis in (A.66).

When $\mu_1(h_1 - h_2) \geq \mu_2 h_2$, a straightforward induction on n based on (A.65) and (A.66) yields $d_{n,\beta}(x_1,x_2) \geq 0$, $x_2 \geq 1$, which combined with $\hat{d}_{n,\beta}(x_1,1) \geq d_{n,\beta}(x_1,1)$ and Proposition 2.3 proves part (ii).

Proof of Theorem 2.11

When $h_1 < h_2$, idling policies may be optimal. Therefore, when there is no dedicated server in Station 2, the optimal allocation is either $(\nu_1+\mu_1,0)$ or $(0,0)$ when there are no jobs in the downstream station, and one of $(\nu_1+\mu_1,0)$, (ν_1,μ_2), and $(0,\mu_2)$ with jobs in both stations. Then, for $x_2 \geq 1$

$$V_{n,\beta}(0,x_2) = h_2 x_2 + \beta [\lambda V_{n-1,\beta}(1,x_2) + \mu_2 V_{n-1,\beta}(0,x_2-1) + (\mu_1+\nu_1) V_{n-1,\beta}(0,x_2)], \tag{A.67}$$

and for $x_1 \geq 1$

$$\begin{aligned} V_{n,\beta}(x_1,0) &= h_1 x_1 + \beta [\lambda V_{n-1,\beta}(x_1+1,0) + \mu_2 V_{n-1,\beta}(x_1,0) \\ &\quad + (\nu_1+\mu_1) \min\{V_{n-1,\beta}(x_1-1,1), V_{n-1,\beta}(x_1,0)\}], \quad \text{(A.68)} \\ V_{n,\beta}(x_1,x_2) &= h_1 x_1 + h_2 x_2 + \beta [\lambda V_{n-1,\beta}(x_1+1,x_2) \\ &\quad + \min\{(\nu_1+\mu_1) V_{n-1,\beta}(x_1-1,x_2+1) + \mu_2 V_{n-1,\beta}(x_1,x_2), \\ &\qquad \nu_1 V_{n-1,\beta}(x_1-1,x_2+1) + \mu_2 V_{n-1,\beta}(x_1,x_2-1) + \mu_1 V_{n-1,\beta}(x_1,x_2), \\ &\qquad \mu_2 V_{n-1,\beta}(x_1,x_2-1) + (\nu_1+\mu_1) V_{n-1,\beta}(x_1,x_2)\}], \quad x_2 \geq 1. \quad \text{(A.69)} \end{aligned}$$

To prove that the optimal policy assigns the flexible server to Station 2 it suffices to show for each time n that allocation (ν_1,μ_2) incurs no more cost than $(\nu_1+\mu_1,0)$, which in view of (A.69) leads to $d_{n,\beta}(x_1,x_2) \leq 0$ for $x_2 \geq 1$. The proof is by induction on n. Assuming that $d_{n-1,\beta}(x_1,x_2) \leq 0$ (induction hypothesis) means that the optimal allocation at time n

for $x_2 \geq 1$ is either (ν_1, μ_2) or $(0, \mu_2)$, so the first term in braces in (A.69) can be omitted. Then, we use (A.67)-(A.69) to get

$$\begin{aligned} d_{n,\beta}(1, x_2) &= \mu_1(h_1 - h_2) - \mu_2 h_2 + \beta \left[\lambda d_{n-1,\beta}(2, x_2) + \mu_1 d_{n-1,\beta}(1, x_2)\right. \\ &\quad + \mu_1 \nu_1 f_{n-1,\beta}(1, x_2)^- - \mu_2 \nu_1 f_{n-1,\beta}(1, x_2 - 1)^+ + \mu_2 \nu_1 \left[g_{n-1,\beta}(1, x_2) + f_{n-1,\beta}(1, x_2)^+\right] \\ &\quad \left. + \mu_2 d_{n-1,\beta}(1, 0)^- \mathbf{1}(x_2 = 1) + \mu_2 d_{n-1,\beta}(1, x_2 - 1)\mathbf{1}(x_2 > 1)\right], \end{aligned} \tag{A.70}$$

and for $x_1 \geq 2$

$$\begin{aligned} d_{n,\beta}(x_1, x_2) &= \mu_1(h_1 - h_2) - \mu_2 h_2 + \beta \left[\lambda d_{n-1,\beta}(x_1 + 1, x_2) + \mu_1 d_{n-1,\beta}(x_1, x_2)\right. \\ &\quad + \nu_1(\mu_1 - \mu_2) f_{n-1,\beta}(x_1, x_2)^- + \nu_1 \left[\mu_2 g_{n-1,\beta}(x_1 - 1, x_2 + 1) + \mu_1 f_{n-1,\beta}(x_1 - 1, x_2 + 1)^+\right] \\ &\quad \left. + \mu_2 d_{n-1,\beta}(x_1, 0)^- \mathbf{1}(x_2 = 1) + \mu_2 d_{n-1,\beta}(x_1, x_2 - 1)\mathbf{1}(x_2 > 1)\right]. \end{aligned} \tag{A.71}$$

Note that for $x_1, x_2 \geq 1$ we have

$$\mu_2 g_{n-1,\beta}(x_1, x_2) + \mu_1 f_{n-1,\beta}(x_1, x_2)^+ = \max\left\{\mu_2 g_{n-1,\beta}(x_1, x_2), d_{n-1,\beta}(x_1, x_2)\right\}. \tag{A.72}$$

Then, the result follows from (A.70)-(A.72), $\mu_1 \geq \mu_2$, and the induction hypothesis.

Appendix B

Proof of Lemma 2.10

We start with the proof of parts (i)-(iii). For $x_2 \geq 1$ we have

$$V(x_1, i_1, j, x_2, 0) \leq V(x_1, i_1, j, x_2 - 1, 1), \tag{B.1}$$
$$V(x_1, i_1, 0, x_2, i_2) \leq V(x_1, i_1, 2, x_2 - 1, i_2), \tag{B.2}$$

because assigning the dedicated server of Station 2 (respectively, the slow server) to Station 2 may not be optimal. Next, consider the following sample path argument. Let $P1$ and $P2$ be the processes that start in states (x_1, i_1, j, x_2, i_2) and $(x_1, i_1, j, x_2+1, i_2)$, respectively, and π be the optimal policy for $P2$. Assume that for $P1$ we apply a policy $\tilde{\pi}$ that imitates π until the first time, say τ, that it is unable to do so, and is optimal afterwards. Policy $\tilde{\pi}$ cannot imitate π only when there is one job in Station 2 under $P2$ and a server is assigned to it, in which case $\tilde{\pi}$ cannot replicate the action because of lack of jobs under $P1$. Then, the state of the system under $P1$ and $P2$ at time τ can be either $(\tilde{x}_1, \tilde{i}_1, \tilde{j}, 0, 0)$ and $(\tilde{x}_1, \tilde{i}_1, \tilde{j}, 0, 1)$, or $(\tilde{x}_1, \tilde{i}_1, 0, 0, \tilde{i}_2)$ and $(\tilde{x}_1, \tilde{i}_1, 2, 0, \tilde{i}_2)$. The two policies have a holding cost rate difference of h_2 until time τ and are optimal afterwards. Therefore, because $\tilde{\pi}$ is not necessarily optimal we have

$$\begin{aligned} & V(x_1, i_1, j, x_2, i_2) - V(x_1, i_1, j, x_2+1, i_2) \\ & \leq -h_2 E(\tau) + E\left[V(\tilde{x}_1, \tilde{i}_1, \tilde{j}, 0, 0) - V(\tilde{x}_1, \tilde{i}_1, \tilde{j}, 0, 1)\right] \end{aligned} \tag{B.3}$$

or

$$\begin{aligned} & V(x_1, i_1, j, x_2, i_2) - V(x_1, i_1, j, x_2+1, i_2) \\ & \leq -h_2 E(\tau) + E\left[V(\tilde{x}_1, \tilde{i}_1, 0, 0, \tilde{i}_2) - V(\tilde{x}_1, \tilde{i}_1, 2, 0, \tilde{i}_2)\right]. \end{aligned} \tag{B.4}$$

Let $X_1 = x_1 + i_1 + \mathbf{1}(j = 1)$ be the number of jobs in Station 1, waiting and in service. We claim that, assuming that parts (i)-(iii) hold for $X_1 \leq K$, it suffices to prove parts (ii) and (iii) for $X_1 = K + 1$ and $x_2 = 0$ for parts (i)-(iii) to hold for $X_1 = K + 1$. This is true because part (i) would follow from (B.3) or (B.4), and then parts (ii) and (iii) for $x_2 \geq 1$ would follow from (B.1) and (B.2). Based on this observation we use induction on X_1 to prove parts (i)-(iii).

The induction base is established by proving parts (ii) and (iii) for $X_1 = x_2 = 0$. For $j = 0$ or $j = 2$ we have

$$V(0, 0, j, 0, 1) - V(0, 0, j, 0, 0) = h_2/\nu_2,$$

and for $i_2 = 0$ or $i_2 = 1$,

$$V(0,0,2,0,i_2) - V(0,0,0,0,i_2) = h_2/\mu_2.$$

The induction step for part (ii) requires the proof of $V(x_1,i_1,j,0,0) \leq V(x_1,i_1,j,0,1)$. We consider all possible combinations of i_1, j and use (2.4.1)-(2.4.4) to derive expressions for $V(x_1,i_1,j,0,0)$ and $V(x_1,i_1,j,0,1)$.
Case 1. $i_1 = j = 1$. We have

$$\begin{aligned}V(x_1,1,1,0,1) - V(x_1,1,1,0,0) = W(x_1,1,1,0,1) - W(x_1,1,1,0,0) = h_2 \\ +\nu_1[V(x_1,0,1,1,1) - V(x_1,0,1,1,0)] + \mu_1[V(x_1,1,0,1,1) - V(x_1,1,0,1,0)] \\ +\mu_2[V(x_1,1,1,0,1) - V(x_1,1,1,0,0)],\end{aligned}$$

and using the induction hypothesis we get

$$V(x_1,1,1,0,0) \leq V(x_1,1,1,0,1). \tag{B.5}$$

Case 2. $i_1 = 0$, $j = 1$. The optimal action of the dedicated server of Station 1 for state $(x_1,0,1,0,1)$, denoted by $\tilde{\alpha}_1$, is feasible for state $(x_1,0,1,0,0)$, so we have

$$V(x_1,0,1,0,1) - V(x_1,0,1,0,0) \geq W(x_1-\tilde{\alpha}_1,\tilde{\alpha}_1,1,0,1) - W(x_1-\tilde{\alpha}_1,\tilde{\alpha}_1,1,0,0). \tag{B.6}$$

For $\tilde{\alpha}_1 = 0$ we have

$$\begin{aligned}W(x_1,0,1,0,1) - W(x_1,0,1,0,0) = h_2 \\ +\mu_1[V(x_1,0,0,1,1) - V(x_1,0,0,1,0)] \\ +(\nu_1+\mu_2)[V(x_1,0,1,0,1) - V(x_1,0,1,0,0)].\end{aligned} \tag{B.7}$$

For $\tilde{\alpha}_1 = 1$ we have

$$W(x_1-1,1,1,0,1) - W(x_1-1,1,1,0,0) = V(x_1-1,1,1,0,1) - V(x_1-1,1,1,0,0) \geq 0 \tag{B.8}$$

from (B.5). Using (B.6),(B.7), and the induction hypothesis for $\tilde{\alpha}_1 = 0$ and (B.6),(B.8) for $\tilde{\alpha}_1 = 1$, we get

$$V(x_1,0,1,0,0) \leq V(x_1,0,1,0,1). \tag{B.9}$$

Case 3. $i_1 = 1$, $j = 0$. Interchanging the role of i_1, j in the analysis of Case 2 we obtain

$$V(x_1,1,0,0,0) \leq V(x_1,1,0,0,1). \tag{B.10}$$

Case 4. $i_1 = j = 0$. Let $\tilde{\alpha}_1, \tilde{\alpha}$ be the optimal actions of the dedicated server of Station 1 and the slow server for state $(x_1,0,0,0,1)$. If $\tilde{\alpha}_1 = \tilde{\alpha} = 0$ we have $V(x_1,0,0,0,1) - V(x_1,0,0,0,0) = h_2/\nu_2$. In any other case actions $\tilde{\alpha}_1, \tilde{\alpha}$ are feasible for state $(x_1,0,0,0,0)$, so we have

$$V(x_1,0,0,0,1) - V(x_1,0,0,0,0) \geq W(x_1-\tilde{\alpha}_1-\tilde{\alpha},\tilde{\alpha}_1,\tilde{\alpha},0,1) - W(x_1-\tilde{\alpha}_1-\tilde{\alpha},\tilde{\alpha}_1,\tilde{\alpha},0,0). \tag{B.11}$$

For $\tilde{\alpha}_1 = 1$, $\tilde{\alpha} = 0$ we have

$$\begin{aligned} &W(x_1-1,1,0,0,1) - W(x_1-1,1,0,0,0) \\ &\quad = h_2 + \nu_1[V(x_1-1,0,0,1,1) - V(x_1-1,0,0,1,0)] \\ &\quad +(\mu_1+\mu_2)[V(x_1-1,1,0,0,1) - V(x_1-1,1,0,0,0)] \geq 0, \end{aligned} \tag{B.12}$$

by induction and (B.10). For $\tilde{\alpha}_1 = 0$, $\tilde{\alpha} = 1$ we have

$$\begin{aligned} &W(x_1-1,0,1,0,1) - W(x_1-1,0,1,0,0) \\ &\quad = h_2 + \mu_1[V(x_1-1,0,0,1,1) - V(x_1-1,0,0,1,0)] \\ &\quad +(\nu_1+\mu_2)[V(x_1-1,0,1,0,1) - V(x_1-1,0,1,0,0)] \geq 0, \end{aligned} \tag{B.13}$$

by induction and (B.9). For $\tilde{\alpha}_1 = \tilde{\alpha} = 1$ we have

$$\begin{aligned} &W(x_1-2,1,1,0,1) - W(x_1-2,1,1,0,0) = h_2 \\ &\quad +\nu_1[V(x_1-2,0,1,1,1) - V(x_1-2,0,1,1,0)] \\ &\quad +\mu_1[V(x_1-2,1,0,1,1) - V(x_1-2,1,0,1,0)] \\ &\quad +\mu_2[V(x_1-2,1,1,0,1) - V(x_1-2,1,1,0,0)] \geq 0, \end{aligned} \tag{B.14}$$

by induction and (B.5). Then, get we from (B.11)-(B.14)

$$V(x_1,0,0,0,0) \leq V(x_1,0,0,0,1). \tag{B.15}$$

<u>Case 5.</u> $i_1 = 1$, $j = 2$. We have

$$\begin{aligned} &V(x_1,1,2,0,1) - V(x_1,1,2,0,0) = W(x_1,1,2,0,1) - W(x_1,1,2,0,0) = h_2 \\ &\quad +\nu_1[V(x_1,0,2,1,1) - V(x_1,0,2,1,0)] + \mu_2[V(x_1,1,0,0,1) - V(x_1,1,0,0,0)] \\ &\quad +\mu_1[V(x_1,1,2,0,1) - V(x_1,1,2,0,0)], \end{aligned}$$

and using the induction hypothesis and (B.10) we get

$$V(x_1,1,2,0,0) \leq V(x_1,1,2,0,1). \tag{B.16}$$

<u>Case 6.</u> $i_1 = 0$, $j = 2$. The optimal action of the dedicated server of Station 1 for state $(x_1,0,2,0,1)$, denoted by $\tilde{\alpha}_1$, is feasible for state $(x_1,0,2,0,0)$, so we have

$$V(x_1,0,2,0,1) - V(x_1,0,2,0,0) \geq W(x_1-\tilde{\alpha}_1,\tilde{\alpha}_1,2,0,1) - W(x_1-\tilde{\alpha}_1,\tilde{\alpha}_1,2,0,0). \tag{B.17}$$

For $\tilde{\alpha}_1 = 0$ we have

$$\begin{aligned} &W(x_1,0,2,0,1) - W(x_1,0,2,0,0) = h_2 \\ &\quad +\mu_2[V(x_1,0,0,0,1) - V(x_1,0,0,0,0)] \\ &\quad +(\nu_1+\mu_1)[V(x_1,0,2,0,1) - V(x_1,0,2,0,0)]. \end{aligned} \tag{B.18}$$

For $\tilde{\alpha}_1 = 1$ we have

$$W(x_1-1,1,2,0,1) - W(x_1-1,1,2,0,0) = V(x_1-1,1,2,0,1) - V(x_1-1,1,2,0,0) \geq 0 \tag{B.19}$$

from (B.16). Using (B.17),(B.18),(B.15) for $\tilde{\alpha}_1 = 0$ and (B.17),(B.19) for $\tilde{\alpha}_1 = 1$, we get

$$V(x_1, 0, 2, 0, 0) \leq V(x_1, 0, 2, 0, 1). \quad \text{(B.20)}$$

The induction step for part (iii) requires the proof of $V(x_1, i_1, 0, 0, i_2) \leq V(x_1, i_1, 2, 0, i_2)$. For every combination of i_1, i_2 we use (2.4.1)-(2.4.4) to derive expressions for $V(x_1, i_1, 0, 0, i_2)$ and $V(x_1, i_1, 2, 0, i_2)$.
Case 1. $i_1 = 1$, $i_2 = 0$. We have

$$\begin{aligned} &V(x_1, 1, 2, 0, 0) - V(x_1, 1, 0, 0, 0) \geq W(x_1, 1, 2, 0, 0) - W(x_1, 1, 0, 0, 0) = h_2 \\ &\quad +\nu_1[V(x_1, 0, 2, 1, 0) - V(x_1, 0, 0, 1, 0)] + (\mu_1 + \nu_2)[V(x_1, 1, 2, 0, 0) - V(x_1, 1, 0, 0, 0)], \end{aligned}$$

and using the induction hypothesis we get

$$V(x_1, 1, 0, 0, 0) \leq V(x_1, 1, 2, 0, 0). \quad \text{(B.21)}$$

Case 2. $i_1 = i_2 = 1$. We have

$$\begin{aligned} &V(x_1, 1, 2, 0, 1) - V(x_1, 1, 0, 0, 1) \geq W(x_1, 1, 2, 0, 1) - W(x_1, 1, 0, 0, 1) = h_2 \\ &\quad +\nu_1[V(x_1, 0, 2, 1, 1) - V(x_1, 0, 0, 1, 1)] + \nu_2[V(x_1, 1, 2, 0, 0) - V(x_1, 1, 0, 0, 0)] \\ &\quad +\mu_1[V(x_1, 1, 2, 0, 1) - V(x_1, 1, 0, 0, 1)], \end{aligned}$$

and using the induction hypothesis and (B.21) we get

$$V(x_1, 1, 0, 0, 1) \leq V(x_1, 1, 2, 0, 1). \quad \text{(B.22)}$$

Case 3. $i_1 = i_2 = 0$. Let $\tilde{\alpha}_1$ be the optimal action of the dedicated server of Station 1 for state $(x_1, 0, 2, 0, 0)$. For $\tilde{\alpha}_1 = 0$ we have

$$V(x_1, 0, 2, 0, 0) = W(x_1, 0, 2, 0, 0) = \frac{h_1 x_1 + h_2}{\mu_2} + V(x_1, 0, 0, 0, 0). \quad \text{(B.23)}$$

For $\tilde{\alpha}_1 = 1$ we have

$$\begin{aligned} &V(x_1, 0, 2, 0, 0) - V(x_1, 0, 0, 0, 0) \\ &\quad \geq W(x_1 - 1, 1, 2, 0, 0) - W(x_1 - 1, 1, 0, 0, 0) = h_2 \\ &\quad +\nu_1[V(x_1 - 1, 0, 2, 1, 0) - V(x_1 - 1, 0, 0, 1, 0)] \\ &\quad +(\nu_2 + \mu_1)[V(x_1 - 1, 1, 2, 0, 0) - V(x_1 - 1, 1, 0, 0, 0)], \end{aligned} \quad \text{(B.24)}$$

where the inequality is due to the fact that actions $1, 0$ for the dedicated server of Station 1 and the slow server, respectively, are feasible for state $(x_1, 0, 0, 0, 0)$. Using (B.23) for $\tilde{\alpha}_1 = 0$ and (B.24),(B.21) and the induction hypothesis for $\tilde{\alpha}_1 = 1$, we get

$$V(x_1, 0, 0, 0, 0) \leq V(x_1, 0, 2, 0, 0). \quad \text{(B.25)}$$

Case 4. $i_1 = 0$, $i_2 = 1$. Let $\tilde{\alpha}_1$ be the optimal action of the dedicated server of Station 1 for state $(x_1, 0, 2, 0, 1)$. Then, actions $\tilde{\alpha}_1, 0$ for the dedicated server of Station 1 and the slow server, respectively, are feasible for state $(x_1, 0, 0, 0, 1)$, so we have

$$V(x_1, 0, 2, 0, 1) - V(x_1, 0, 0, 0, 1) \geq W(x_1 - \tilde{\alpha}_1, \tilde{\alpha}_1, 2, 0, 1) - W(x_1 - \tilde{\alpha}_1, \tilde{\alpha}_1, 0, 0, 1). \quad \text{(B.26)}$$

For $\tilde{\alpha}_1 = 0$ we have

$$\begin{aligned}W(x_1, 0, 2, 0, 1) - W(x_1, 0, 0, 0, 1) = h_2 \\ +\nu_2[V(x_1, 0, 2, 0, 0) - V(x_1, 0, 0, 0, 0)] \\ +(\nu_1 + \mu_1)[V(x_1, 0, 2, 0, 1) - V(x_1, 0, 0, 0, 1)].\end{aligned} \tag{B.27}$$

For $\tilde{\alpha}_1 = 1$ we have

$$\begin{aligned}W(x_1 - 1, 1, 2, 0, 1) - W(x_1 - 1, 1, 0, 0, 1) = h_2 \\ +\nu_1[V(x_1 - 1, 0, 2, 1, 1) - V(x_1 - 1, 0, 0, 1, 1)] \\ +\nu_2[V(x_1 - 1, 1, 2, 0, 0) - V(x_1 - 1, 1, 0, 0, 0)] \\ +\mu_1[V(x_1 - 1, 1, 2, 0, 1) - V(x_1 - 1, 1, 0, 0, 1)].\end{aligned} \tag{B.28}$$

Using (B.26),(B.27),(B.25) for $\tilde{\alpha}_1 = 0$ and (B.26),(B.28),(B.21),(B.22), and the induction hypothesis for $\tilde{\alpha}_1 = 1$, we get

$$V(x_1, 0, 0, 0, 1) \leq V(x_1, 0, 2, 0, 1). \tag{B.29}$$

We now proceed to the proof of parts (iv) and (v). For $x_2 \geq 1$ we have

$$V(x_1, i_1, 0, x_2, 1) \leq V(x_1, i_1, 2, x_2 - 1, 1), \tag{B.30}$$

because assigning the slow server to Station 2 may not be optimal. Next, we use a sample path argument to compare the terms involved in part (iv). Let $P1$ and $P2$ be the processes that start in states $(x_1, i_1, j, x_2 - 1, 1)$ and $(x_1, i_1, j, x_2, 0)$, respectively, and π be the optimal policy for $P2$. Assume that for $P1$ we apply a policy $\tilde{\pi}$ that imitates π until time τ (defined later), and is optimal afterwards. Time τ is the time that the earliest of the following three events occurs: i) Server 2 is assigned to a job under $P2$, ii) a service completion by Server 2 under $P1$, and iii) there is one job in Station 2 under $P2$ and the slow server is assigned to it, in which case $\tilde{\pi}$ cannot replicate this action because of lack of jobs under $P1$. In the first case the two processes are coupled and there is no cost difference between π and $\tilde{\pi}$. In the second case the state of the system under $P1$ and $P2$ at time τ is of the form $(\tilde{x}_1, \tilde{i}_1, \tilde{j}, \tilde{x}_2 - 1, 0)$ and $(\tilde{x}_1, \tilde{i}_1, \tilde{j}, \tilde{x}_2, 0)$, respectively. Finally, in the third case the state of the system under $P1$ and $P2$ at time τ is of the form $(\tilde{x}_1, \tilde{i}_1, 0, 0, 1)$ and $(\tilde{x}_1, \tilde{i}_1, 2, 0, 0)$. Therefore, because $\tilde{\pi}$ is not necessarily optimal we have

$$V(x_1, i_1, j, x_2 - 1, 1) - V(x_1, i_1, j, x_2, 0) \leq 0, \tag{B.31}$$

or

$$V(x_1, i_1, j, x_2 - 1, 1) - V(x_1, i_1, j, x_2, 0) \leq E\left[V(\tilde{x}_1, \tilde{i}_1, \tilde{j}, \tilde{x}_2 - 1, 0) - V(\tilde{x}_1, \tilde{i}_1, \tilde{j}, \tilde{x}_2, 0)\right], \tag{B.32}$$

or

$$V(x_1, i_1, j, x_2 - 1, 1) - V(x_1, i_1, j, x_2, 0) \leq E\left[V(\tilde{x}_1, \tilde{i}_1, 0, 0, 1) - V(\tilde{x}_1, \tilde{i}_1, 2, 0, 0)\right]. \tag{B.33}$$

Let $X_1 = x_1 + i_1 + \mathbf{1}(j = 1)$ be the number of jobs in Station 1, waiting and in service. We claim that, assuming that parts (iv) and (v) hold for $X_1 \leq K$, it suffices to prove part (v) for $X_1 = K+1$ and $x_2 = 0$ for parts (iv) and (v) to hold for $X_1 = K+1$. This is true because part (iv) would follow from (B.31) or (B.32) (part (i) of the lemma) or (B.33), and then part (v) for $x_2 \geq 1$ would follow from (B.30). Based on this observation we use induction on X_1 to prove parts (iv) and (v).

The induction base is established by proving part (v) for $X_1 = x_2 = 0$. We have

$$V(0,0,2,0,0) - V(0,0,0,0,1) = \frac{h_2}{\mu_2} - \frac{h_2}{\nu_2} \geq 0,$$

because $\nu_2 > \mu_2$. For the induction step we need to prove $V(x_1, i_1, 0, 0, 1) \leq V(x_1, i_1, 2, 0, 0)$.
Case 1. $i_1 = 1$. We have

$$\begin{aligned}
&V(x_1,1,2,0,0) - V(x_1,1,0,0,1) \geq W(x_1,1,2,0,0) - W(x_1,1,0,0,1) \\
&= \nu_1[V(x_1,0,2,1,0) - V(x_1,0,0,1,1)] + \nu_2[V(x_1,1,2,0,0) - V(x_1,1,0,0,0)] \\
&+\mu_1[V(x_1,1,2,0,0) - V(x_1,1,0,0,1)] + \mu_2[V(x_1,1,0,0,0) - V(x_1,1,0,0,1)].
\end{aligned}$$

Because the term multiplying ν_2 is nonnegative (part (iii) of the lemma) and $\nu_2 > \mu_2$ we get

$$\begin{aligned}
&V(x_1,1,2,0,0) - V(x_1,1,0,0,1) \\
&\quad \geq \nu_1[V(x_1,0,2,1,0) - V(x_1,0,0,1,1)] \\
&\quad +(\mu_1+\mu_2)[V(x_1,1,2,0,0) - V(x_1,1,0,0,0)],
\end{aligned}$$

and using the induction hypothesis we get

$$V(x_1,1,0,0,1) \leq V(x_1,1,2,0,0). \tag{B.34}$$

Case 2. $i_1 = 0$. Let $\tilde{\alpha}_1$ be the optimal action of the dedicated server of Station 1 for state $(x_1,0,2,0,0)$. Then, actions $\tilde{\alpha}_1, 0$ for the dedicated server of Station 1 and the slow server, respectively, are feasible for state $(x_1,0,0,0,1)$, so we have

$$V(x_1,0,2,0,0) - V(x_1,0,0,0,1) \geq W(x_1 - \tilde{\alpha}_1, \tilde{\alpha}_1, 2,0,0) - W(x_1 - \tilde{\alpha}_1, \tilde{\alpha}_1, 0,0,1). \tag{B.35}$$

For $\tilde{\alpha}_1 = 0$ we have

$$W(x_1,0,2,0,0) - W(x_1,0,0,0,1) = \frac{h_1 x_1 + h_2}{\mu_2} - \frac{h_1 x_1 + h_2}{\nu_2} \geq 0, \tag{B.36}$$

because $\nu_2 > \mu_2$. For $\tilde{\alpha}_1 = 1$ we have

$$\begin{aligned}
&W(x_1-1,1,2,0,0) - W(x_1-1,1,0,0,1) \\
&\quad = \nu_1[V(x_1-1,0,2,1,0) - V(x_1-1,0,0,1,1)] \\
&\quad +\nu_2[V(x_1-1,1,2,0,0) - V(x_1-1,1,0,0,0)] \\
&\quad +\mu_1[V(x_1-1,1,2,0,0) - V(x_1-1,1,0,0,1)] \\
&\quad +\mu_2[V(x_1-1,1,0,0,0) - V(x_1-1,1,0,0,1)].
\end{aligned}$$

Because the term multiplying ν_2 is nonnegative (part (iii) of the lemma) and $\nu_2 > \mu_2$ we get

$$\begin{aligned} &W(x_1-1,1,2,0,0) - W(x_1-1,1,0,0,1) \\ &\quad \geq \nu_1[V(x_1-1,0,2,1,0) - V(x_1-1,0,0,1,1)] \\ &\quad +(\mu_1+\mu_2)[V(x_1-1,1,2,0,0) - V(x_1-1,1,0,0,1)]. \qquad \text{(B.37)} \end{aligned}$$

Using (B.35),(B.36) for $\tilde{\alpha}_1 = 0$ and (B.35),(B.37),(B.34), and the induction hypothesis for $\tilde{\alpha}_1 = 1$, we get

$$V(x_1,0,0,0,1) \leq V(x_1,0,2,0,0),$$

completing the proof.

Proof of Lemma 2.11

We start with the proof of parts (i)-(iii). For $x_1 \geq 1$ we have

$$V(x_1,0,j,x_2) \leq V(x_1-1,1,j,x_2), \qquad \text{(B.38)}$$

$$V(x_1,i_1,0,x_2) \leq V(x_1-1,i_1,1,x_2), \qquad \text{(B.39)}$$

because assigning the dedicated server of Station 1 (respectively, the slow server) to Station 1 may not be optimal. Next, we adjust the sample path argument used in the proof of parts (i)-(iii) of Lemma 2.10 to compare the expected cost of processes $P1$ and $P2$ that start in states (x_1,i_1,j,x_2) and (x_1+1,i_1,j,x_2), respectively, by defining policies π and $\tilde{\pi}$ accordingly. The two policies have a holding cost rate difference of h_1 until time τ, the first time that $\tilde{\pi}$ cannot imitate π, and the state of the system under $P1$ and $P2$ at that time can be either $(0,0,\tilde{j},\tilde{x}_2)$ and $(0,1,\tilde{j},\tilde{x}_2)$, or $(0,\tilde{i}_1,0,\tilde{x}_2)$ and $(0,\tilde{i}_1,1,\tilde{x}_2)$. Therefore,

$$V(x_1,i_1,j,x_2) - V(x_1+1,i_1,j,x_2) \leq -h_1E(\tau) + E\left[V(0,0,\tilde{j},\tilde{x}_2) - V(0,1,\tilde{j},\tilde{x}_2)\right] \qquad \text{(B.40)}$$

or

$$V(x_1,i_1,j,x_2) - V(x_1+1,i_1,j,x_2) \leq -h_1E(\tau) + E\left[V(0,\tilde{i}_1,0,\tilde{x}_2) - V(0,\tilde{i}_1,1,\tilde{x}_2)\right]. \qquad \text{(B.41)}$$

Then, it is evident from (B.38)-(B.41) that it suffices to prove parts (ii) and (iii) for $x_1 = 0$. The proof is by induction on x_2. We start with part (ii) for $j = 0, 2$. For $j = 0$ the induction base is established by (2.4.12). For $j = 2$ we have

$$\begin{aligned} &V(0,1,2,0) - V(0,0,2,0) = W(0,1,2,0) - W(0,0,2,0) = h_1 \\ &\quad +\nu_1[V(0,0,2,1) - V(0,0,2,0)] + \mu_2[V(0,1,0,0) - V(0,0,0,0)] \\ &\quad +(\mu_1+\nu_2)[V(0,1,2,0) - V(0,0,2,0)], \end{aligned}$$

which combined with (2.4.13) and (2.4.12) yields $V(0,0,2,0) \leq V(0,1,2,0)$. For the induction step we need to prove part (ii) for $j = 0, 2$ and $x_2 > 0$. We consider $j = 0$ first. Let $\tilde{\alpha}$ be the optimal action of the slow server for state $(0,1,0,x_2)$. Because $\tilde{\alpha}$ is feasible for state $(0,0,0,x_2)$ we have

$$V(0,1,0,x_2) - V(0,0,0,x_2) \geq W(0,1,\tilde{\alpha},x_2 - \mathbf{1}(\tilde{\alpha}=2)) - W(0,0,\tilde{\alpha},x_2 - \mathbf{1}(\tilde{\alpha}=2)). \qquad \text{(B.42)}$$

For $\tilde{\alpha} = 0$ we have

$$\begin{aligned} W(0,1,0,x_2) - W(0,0,0,x_2) = h_1 \\ +\nu_1[V(0,0,0,x_2+1) - V(0,0,0,x_2)] \\ +\nu_2[V(0,1,0,x_2-1) - V(0,0,0,x_2-1)] \\ +(\mu_1+\mu_2)[V(0,1,0,x_2) - V(0,0,0,x_2)]. \end{aligned} \tag{B.43}$$

For $\tilde{\alpha} = 2$ we have

$$\begin{aligned} W(0,1,2,x_2-1) - W(0,0,2,x_2-1) = h_1 \\ +\nu_1[V(0,0,2,x_2) - V(0,0,2,x_2-1)] \\ +\nu_2[V(0,1,2,x_2-2) - V(0,0,2,x_2-2)] \\ +\mu_1[V(0,1,2,x_2-1) - V(0,0,2,x_2-1)] \\ +\mu_2[V(0,1,0,x_2-1) - V(0,0,0,x_2-1)]. \end{aligned} \tag{B.44}$$

Using (B.42)-(B.44),(2.4.13), and the induction hypothesis we get

$$V(0,0,0,x_2) \le V(0,1,0,x_2). \tag{B.45}$$

For $j = 2$ we have

$$\begin{aligned} V(0,1,2,x_2) - V(0,0,2,x_2) = W(0,1,2,x_2) - W(0,0,2,x_2) = h_1 \\ +\nu_1[V(0,0,2,x_2+1) - V(0,0,2,x_2)] + \nu_2[V(0,1,2,x_2-1) - V(0,0,2,x_2-1)] \\ +\mu_1[V(0,1,2,x_2) - V(0,0,2,x_2)] + \mu_2[V(0,1,0,x_2) - V(0,0,0,x_2)], \end{aligned}$$

and using (2.4.13),(B.45), and the induction hypothesis we get

$$V(0,0,2,x_2) \le V(0,1,2,x_2).$$

For any $x_2 \ge 0$ we have

$$\begin{aligned} V(0,1,1,x_2) - V(0,0,1,x_2) = W(0,1,1,x_2) - W(0,0,1,x_2) \\ = h_1 + \nu_1[V(0,0,1,x_2+1) - V(0,0,1,x_2)] \\ +\nu_2[V(0,1,1,(x_2-1)^+) - V(0,0,1,(x_2-1)^+)] \\ +\mu_1[V(0,1,0,x_2+1) - V(0,0,0,x_2+1)] \\ +\mu_2[V(0,1,1,x_2) - V(0,0,1,x_2)]. \end{aligned} \tag{B.46}$$

Then, taking into account (2.4.13) and (B.45), part (ii) for $j = 1$ follows from a straightforward induction based on (B.46). For part (iii) we consider $i_1 = 0$ first, in which case the induction base is established by (2.4.12). For $x_2 \ge 1$ we have

$$\begin{aligned} V(0,0,1,x_2) - V(0,0,0,x_2) \ge W(0,0,1,x_2) - W(0,0,0,x_2) = h_1 \\ +\nu_2[V(0,0,1,x_2-1) - V(0,0,0,x_2-1)] + \mu_1[V(0,0,0,x_2+1) - V(0,0,0,x_2)] \\ +(\nu_1+\mu_2)[V(0,0,1,x_2) - V(0,0,0,x_2)], \end{aligned}$$

and using (2.4.13) and the induction hypothesis we get

$$V(0,0,0,x_2) \leq V(0,0,1,x_2). \tag{B.47}$$

For any $x_2 \geq 0$ we have

$$\begin{aligned} V(0,1,1,x_2) - V(0,1,0,x_2) &\geq W(0,1,1,x_2) - W(0,1,0,x_2) \\ &= h_1 + \nu_1[V(0,0,1,x_2+1) - V(0,0,0,x_2+1)] \\ &+\nu_2[V(0,1,1,(x_2-1)^+) - V(0,1,0,(x_2-1)^+)] \\ &+\mu_1[V(0,1,0,x_2+1) - V(0,1,0,x_2)] \\ &+\mu_2[V(0,1,1,x_2) - V(0,1,0,x_2)]. \end{aligned} \tag{B.48}$$

Then, taking into account (2.4.13) and (B.47), part (iii) for $i_1 = 1$ follows from a straightforward induction based on (B.48).

We now proceed to prove parts (iv)-(vi). Similarly to the proof of (i)-(iii), it suffices to prove parts (v) and (vi) for $x_1 = 0$ because i) (B.38),(B.39) hold with $x_2 + 1$ instead of x_2, and ii) the same sample path argument can be applied to compare the expected cost of the processes that start in states $(x_1 - 1, i_1, j, x_2 + 1)$ and (x_1, i_1, j, x_2), yielding

$$\begin{aligned} &V(x_1-1,i_1,j,x_2+1) - V(x_1,i_1,j,x_2) \\ &\quad \leq -(h_1-h_2)E(\tau) + E\left[V(0,0,\tilde{j},\tilde{x}_2+1) - V(0,1,\tilde{j},\tilde{x}_2)\right] \end{aligned}$$

or

$$\begin{aligned} &V(x_1-1,i_1,j,x_2+1) - V(x_1,i_1,j,x_2) \\ &\quad \leq -(h_1-h_2)E(\tau) + E\left[V(0,\tilde{i}_1,0,\tilde{x}_2+1) - V(0,\tilde{i}_1,1,\tilde{x}_2)\right]. \end{aligned}$$

We start with part (v) for $j = 0, 2$. We have

$$V(0,1,0,0) - V(0,0,0,1) = h_1/\nu_1, \tag{B.49}$$

and

$$\begin{aligned} V(0,1,2,0) - V(0,0,2,1) &= W(0,1,2,0) - W(0,0,2,1) = h_1 - h_2 \\ &+\nu_2[V(0,1,2,0) - V(0,0,2,0)] + \mu_2[V(0,1,0,0) - V(0,0,0,1)] \\ &+\mu_1[V(0,1,2,0) - V(0,0,2,1)], \end{aligned} \tag{B.50}$$

and the induction base ($x_2 = 0$) is established by (B.49),(B.50) and part (ii). For the induction step we need to prove part (v) for $j = 0, 2$ and $x_2 > 0$. We consider $j = 0$ first. Let $\tilde{\alpha}$ be the optimal action of the slow server for state $(0,1,0,x_2)$. Because $\tilde{\alpha}$ is feasible for state $(0,0,0,x_2+1)$ we have

$$V(0,1,0,x_2) - V(0,0,0,x_2+1) \geq W(0,1,\tilde{\alpha},x_2 - \mathbf{1}(\tilde{\alpha}=2)) - W(0,0,\tilde{\alpha},x_2 + \mathbf{1}(\tilde{\alpha}=0)). \tag{B.51}$$

For $\tilde{\alpha} = 0$ we have

$$\begin{aligned} W(0,1,0,x_2) - W(0,0,0,x_2+1) &= h_1 - h_2 \\ &+\nu_2[V(0,1,0,x_2-1) - V(0,0,0,x_2)] \\ &+(\mu_1+\mu_2)[V(0,1,0,x_2) - V(0,0,0,x_2+1)]. \end{aligned} \tag{B.52}$$

For $\tilde{\alpha} = 2$ we have

$$\begin{aligned} W(0,1,2,x_2-1) - W(0,0,2,x_2) &= h_1 - h_2 \\ &+\nu_2[V(0,1,2,x_2-2) - V(0,0,2,x_2-1)] \\ &+\mu_1[V(0,1,2,x_2-1) - V(0,0,2,x_2)] \\ &+\mu_2[V(0,1,0,x_2-1) - V(0,0,0,x_2)]. \end{aligned} \tag{B.53}$$

Using (B.51)-(B.53) and the induction hypothesis we get

$$V(0,0,0,x_2+1) \le V(0,1,0,x_2). \tag{B.54}$$

For $j = 2$ we have

$$\begin{aligned} &V(0,1,2,x_2) - V(0,0,2,x_2+1) = W(0,1,2,x_2) - W(0,0,2,x_2+1) = h_1 - h_2 \\ &\quad +\nu_2[V(0,1,2,x_2-1) - V(0,0,2,x_2)] + \mu_1[V(0,1,2,x_2) - V(0,0,2,x_2+1)] \\ &\quad +\mu_2[V(0,1,0,x_2) - V(0,0,0,x_2+1)], \end{aligned}$$

and using (B.54) and the induction hypothesis we get

$$V(0,0,2,x_2+1) \le V(0,1,2,x_2).$$

For any $x_2 \ge 0$ we have

$$\begin{aligned} V(0,1,1,x_2) - V(0,0,1,x_2+1) &= W(0,1,1,x_2) - W(0,0,1,x_2+1) \\ &= h_1 - h_2 + \nu_2[V(0,1,1,(x_2-1)^+) - V(0,0,1,x_2)] \\ &+\mu_1[V(0,1,0,x_2+1) - V(0,0,0,x_2+2)] \\ &+\mu_2[V(0,1,1,x_2) - V(0,0,1,x_2+1)]. \end{aligned} \tag{B.55}$$

Then, taking into account (B.54), part (v) for $j = 1$ follows from a straightforward induction based on (B.55). For part (vi) we consider $i_1 = 0$ first, in which case the induction base is established by $V(0,0,1,0) - V(0,0,0,1) = h_1/\mu_1$. For $x_2 \ge 1$ we have

$$\begin{aligned} &V(0,0,1,x_2) - V(0,0,0,x_2+1) \ge W(0,0,1,x_2) - W(0,0,0,x_2+1) = h_1 - h_2 \\ &\quad +\nu_2[V(0,0,1,x_2-1) - V(0,0,0,x_2)] + (\nu_1+\mu_2)[V(0,0,1,x_2) - V(0,0,0,x_2+1)], \end{aligned}$$

and by the induction hypothesis we get

$$V(0,0,0,x_2+1) \le V(0,0,1,x_2). \tag{B.56}$$

For any $x_2 \geq 0$ we have

$$\begin{aligned} V(0,1,1,x_2) - V(0,1,0,x_2+1) &\geq W(0,1,1,x_2) - W(0,1,0,x_2+1) \\ &= h_1 - h_2 + \nu_1[V(0,0,1,x_2+1) - V(0,0,0,x_2+2)] \\ &+\nu_2[V(0,1,1,(x_2-1)^+) - V(0,1,0,x_2)] \\ &+\mu_2[V(0,1,1,x_2) - V(0,1,0,x_2+1)]. \end{aligned} \tag{B.57}$$

Then, taking into account (B.56), part (vi) for $i_1 = 1$ follows from a straightforward induction based on (B.57).

Finally, the proof of parts (vii) and (viii) is similar to the proof of parts (iv) and (v) of Lemma 2.10. For $x_1 \geq 1$ we have

$$V(x_1,1,0,x_2) \leq V(x_1-1,1,1,x_2), \tag{B.58}$$

because assigning the slow server to Station 1 may not be optimal. Next, we use a similar sample path argument (interchanging the roles of Stations 1,2 and their dedicated servers) to compare the expected cost of the processes that start in states $(x_1-1,1,j,x_2)$ and $(x_1,0,j,x_2)$. Reasoning as in the proof of Lemma 2.10 we get

$$V(x_1-1,1,j,x_2) - V(x_1,0,j,x_2) \leq 0, \tag{B.59}$$

or

$$V(x_1-1,1,j,x_2) - V(x_1,0,j,x_2) \leq E\left[V(\tilde{x}_1-1,0,\tilde{j},\tilde{x}_2+1) - V(\tilde{x}_1,0,\tilde{j},\tilde{x}_2)\right] \leq 0 \tag{B.60}$$

by part (iv), or

$$V(x_1-1,1,j,x_2) - V(x_1,0,j,x_2) \leq E\left[V(0,1,0,\tilde{x}_2) - V(0,0,1,\tilde{x}_2)\right]. \tag{B.61}$$

Then, it is clear from (B.58)-(B.61) that it suffices to prove part (viii) for $x_1 = 0$. The proof is by induction on x_2, with the induction base established by

$$V(0,0,1,0) - V(0,1,0,0) = \frac{h_1}{\mu_1} - \frac{h_1}{\nu_1} \geq 0,$$

because $\nu_1 > \mu_1$. For $x_2 \geq 1$ we have

$$\begin{aligned} &V(0,0,1,x_2) - V(0,1,0,x_2) \geq W(0,0,1,x_2) - W(0,1,0,x_2) \\ &= \nu_1[V(0,0,1,x_2) - V(0,0,0,x_2+1)] + \nu_2[V(0,0,1,x_2-1) - V(0,1,0,x_2-1)] \\ &+\mu_1[V(0,0,0,x_2+1) - V(0,1,0,x_2)] + \mu_2[V(0,0,1,x_2) - V(0,1,0,x_2)]. \end{aligned}$$

Because the term multiplying ν_1 is nonnegative (part (vi) of the lemma) and $\nu_1 > \mu_1$ we get

$$\begin{aligned} &V(0,0,1,x_2) - V(0,1,0,x_2) \\ &\qquad \geq \nu_2[V(0,0,1,x_2-1) - V(0,1,0,x_2-1)] \\ &\qquad +(\mu_1+\mu_2)[V(0,0,1,x_2) - V(0,1,0,x_2)], \end{aligned}$$

and using the induction hypothesis we get

$$V(0,1,0,x_2) \leq V(0,0,1,x_2),$$

completing the proof.

Proof of properties (P1)-(P5)

The proof is by induction on n. First, we use (2.4.19)-(2.4.21) to show that $W_1(x_1, i_1, j, x_2, i_2)$ satisfies (P1)-(P5), establishing the induction base. Then, the induction is completed by assuming that W_n satisfies (P1)-(P5) and showing that V_n and W_{n+1} satisfy these properties as well. According to the induction hypothesis, if there are jobs in Station 2 at time n and its dedicated server is available, the optimal policy assigns a job to that server, that is, for $x_2 \geq 1$ we have

$$V_n(x_1, i_1, j, x_2, 0) = V_n(x_1, i_1, j, x_2 - 1, 1). \tag{B.62}$$

Starting with the proof of (P1) for V_n, let $(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\alpha})$ be the optimal allocation for state $(x_1, i_1, j, x_2 + 1, i_2)$. Then, if $(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\alpha})$ is feasible for state (x_1, i_1, j, x_2, i_2), we have from (2.4.18)

$$\begin{aligned} &V_n(x_1, i_1, j, x_2, i_2) \leq W_n(x_1', \tilde{\alpha}_1, \tilde{\alpha}, x_2', \tilde{\alpha}_2) \\ &\leq W_n(x_1', \tilde{\alpha}_1, \tilde{\alpha}, x_2' + 1, \tilde{\alpha}_2) = V_n(x_1, i_1, j, x_2 + 1, i_2), \end{aligned}$$

where $x_k' = x_k - \tilde{\alpha}_k(1 - i_k) - \mathbf{1}(j = 0, \tilde{\alpha} = k)$, $k = 1, 2$, and the second inequality follows from property (P1) for W_n. Next, we consider the cases for which $(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\alpha})$ is not feasible for state (x_1, i_1, j, x_2, i_2).

<u>Case 1.</u> $x_2 = 0$, $i_2 = 0$. We have $\tilde{\alpha}_2 = 1$ and $\tilde{\alpha} = j$ if $j \neq 0$, $\tilde{\alpha} \neq 2$ if $j = 0$, so $(\tilde{\alpha}_1, 0, \tilde{\alpha})$ is feasible for state $(x_1, i_1, j, 0, 0)$. Therefore,

$$V_n(x_1, i_1, j, 0, 0) \leq W_n(x_1', \tilde{\alpha}_1, \tilde{\alpha}, 0, 0) \leq W_n(x_1', \tilde{\alpha}_1, \tilde{\alpha}, 0, 1) = V_n(x_1, i_1, j, 1, 0),$$

where $x_1' = x_1 - \tilde{\alpha}_1(1 - i_1) - \tilde{\alpha} \cdot \mathbf{1}(j = 0)$ and the second inequality follows from property (P2) for W_n.

<u>Case 2.</u> $x_2 = 0$, $j = 0$, $i_2 = 1$, $\tilde{\alpha} = 2$. In this case $(\tilde{\alpha}_1, 1, 0)$ is feasible for state $(x_1, i_1, 0, 0, 1)$, so we have

$$V_n(x_1, i_1, 0, 0, 1) \leq W_n(x_1', \tilde{\alpha}_1, 0, 0, 1) \leq W_n(x_1', \tilde{\alpha}_1, 2, 0, 1) = V_n(x_1, i_1, 0, 1, 1),$$

where $x_1' = x_1 - \tilde{\alpha}_1(1 - i_1)$ and the second inequality follows from property (P3) for W_n.

<u>Case 3.</u> $x_2 = 1$, $i_2 = j = 0$, $\tilde{\alpha} = 2$. We have $\tilde{\alpha}_2 = 1$, so $(\tilde{\alpha}_1, 1, 0)$ is feasible for state $(x_1, i_1, 0, 1, 0)$. Therefore,

$$V_n(x_1, i_1, 0, 1, 0) \leq W_n(x_1', \tilde{\alpha}_1, 0, 0, 1) \leq W_n(x_1', \tilde{\alpha}_1, 2, 0, 1) = V_n(x_1, i_1, 0, 2, 0),$$

where $x_1' = x_1 - \tilde{\alpha}_1(1 - i_1)$ and the second inequality follows from property (P3) for W_n.

Having completed the proof of property (P1), we also use (B.62) to get

$$V_n(x_1, i_1, j, x_2, 0) \leq V_n(x_1, i_1, j, x_2 + 1, 0) = V_n(x_1, i_1, j, x_2, 1),$$

which proves property (P2). To prove (P3), letting $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$ be the optimal allocations of the dedicated servers for state $(x_1, i_1, 2, x_2, i_2)$, we note that allocation $(\tilde{\alpha}_1, \tilde{\alpha}_2, 0)$ is feasible for state $(x_1, i_1, 0, x_2, i_2)$. Therefore,

$$V_n(x_1, i_1, 0, x_2, i_2) \leq W_n(x_1', \tilde{\alpha}_1, 0, x_2', \tilde{\alpha}_2) \leq W_n(x_1', \tilde{\alpha}_1, 2, x_2', \tilde{\alpha}_2) = V_n(x_1, i_1, 2, x_2, i_2),$$

where $x_k' = x_k - \tilde{\alpha}_k(1 - i_k)$, $k = 1, 2$, and the second inequality follows from property (P3) for W_n.

As a consequence of the induction hypothesis, property (P4) for V_n is satisfied with equality (see (B.62)). We also use the induction hypothesis to prove (P5) for $x_2 \geq 1$, in which case we get

$$V_n(x_1, i_1, 0, x_2, 1) \leq V_n(x_1, i_1, 2, x_2 - 1, 1) = V_n(x_1, i_1, 2, x_2, 0),$$

where the inequality is due to the fact that assigning the flexible server to station 2 may not be optimal. For $x_2 = 0$, we let $\tilde{\alpha}_1$ be the optimal allocation of the dedicated server of Station 1 for state $(x_1, i_1, 2, 0, 0)$. Then, allocation $(\tilde{\alpha}_1, 1, 0)$ is feasible for state $(x_1, i_1, 0, 0, 1)$ and we obtain

$$V_n(x_1, i_1, 0, 0, 1) \leq W_n(x_1', \tilde{\alpha}_1, 0, 0, 1) \leq W_n(x_1', \tilde{\alpha}_1, 2, 0, 0) = V_n(x_1, i_1, 2, 0, 0),$$

where $x_1' = x_1 - \tilde{\alpha}_1(1 - i_1)$ and the second inequality follows from property (P5) for W_n.

Finally, we show that W_{n+1} satisfies properties (P1)-(P5) by using (2.4.19) and (2.4.20). For property (P1) we have

$$\begin{aligned}
&W_{n+1}(x_1, i_1, j, x_2, i_2) - W_{n+1}(x_1, i_1, j, x_2 + 1, i_2) = -h_2 \\
&\quad +\beta \{\lambda[V_n(x_1 + 1, i_1, j, x_2, i_2) - V_n(x_1 + 1, i_1, j, x_2 + 1, i_2)] \\
&\quad +\nu_1[V_n(x_1, 0, j, x_2 + i_1, i_2) - V_n(x_1, 0, j, x_2 + 1 + i_1, i_2)] \\
&\quad +\nu_2[V_n(x_1, i_1, j, x_2, 0) - V_n(x_1, i_1, j, x_2 + 1, 0)] \\
&\quad +\mu_1[V_n(x_1, i_1, j \cdot \mathbf{1}(j \neq 1), x_2 + \mathbf{1}(j = 1), i_2) \\
&\quad -V_n(x_1, i_1, j \cdot \mathbf{1}(j \neq 1), x_2 + 1 + \mathbf{1}(j = 1), i_2)] \\
&\quad +\mu_2[V_n(x_1, i_1, \mathbf{1}(j = 1), x_2, i_2) - V_n(x_1, i_1, \mathbf{1}(j = 1), x_2 + 1, i_2)]\} < 0,
\end{aligned}$$

because V_n satisfies (P1). For property (P2) we have

$$\begin{aligned}
&W_{n+1}(x_1, i_1, j, x_2, 0) - W_{n+1}(x_1, i_1, j, x_2, 1) = -h_2 \\
&\quad +\beta\{\lambda[V_n(x_1 + 1, i_1, j, x_2, 0) - V_n(x_1 + 1, i_1, j, x_2, 1)] \\
&\quad +\nu_1[V_n(x_1, 0, j, x_2 + i_1, 0) - V_n(x_1, 0, j, x_2 + i_1, 1)] \\
&\quad +\mu_1[V_n(x_1, i_1, j \cdot \mathbf{1}(j \neq 1), x_2 + \mathbf{1}(j = 1), 0) \\
&\quad -V_n(x_1, i_1, j \cdot \mathbf{1}(j \neq 1), x_2 + \mathbf{1}(j = 1), 1)] \\
&\quad +\mu_2[V_n(x_1, i_1, \mathbf{1}(j = 1), x_2, 0) - V_n(x_1, i_1, \mathbf{1}(j = 1), x_2, 1)]\} < 0,
\end{aligned}$$

because V_n satisfies (P2). For property (P3) we have

$$\begin{aligned}
&W_{n+1}(x_1, i_1, 0, x_2, i_2) - W_{n+1}(x_1, i_1, 2, x_2, i_2) = -h_2 \\
&\quad +\beta\{\lambda[V_n(x_1 + 1, i_1, 0, x_2, i_2) - V_n(x_1 + 1, i_1, 2, x_2, i_2)] \\
&\quad +\nu_1[V_n(x_1, 0, 0, x_2 + i_1, i_2) - V_n(x_1, 0, 2, x_2 + i_1, i_2)] \\
&\quad +\nu_2[V_n(x_1, i_1, 0, x_2, 0) - V_n(x_1, i_1, 2, x_2, 0)] \\
&\quad +\mu_1[V_n(x_1, i_1, 0, x_2, i_2) - V_n(x_1, i_1, 2, x_2, i_2)]\} < 0,
\end{aligned}$$

because V_n satisfies (P3). For property (P4) we have

$$\begin{aligned}
&W_{n+1}(x_1, i_1, j, x_2 - 1, 1) - W_{n+1}(x_1, i_1, j, x_2, 0)\\
&\quad = \beta\{\lambda[V_n(x_1 + 1, i_1, j, x_2 - 1, 1) - V_n(x_1 + 1, i_1, j, x_2, 0)]\\
&\quad +\nu_1[V_n(x_1, 0, j, x_2 - 1 + i_1, 1) - V_n(x_1, 0, j, x_2 + i_1, 0)]\\
&\quad +\nu_2[V_n(x_1, i_1, j, x_2 - 1, 0) - V_n(x_1, i_1, j, x_2, 0)]\\
&\quad +\mu_1[V_n(x_1, i_1, j \cdot \mathbf{1}(j \neq 1), x_2 - 1 + \mathbf{1}(j = 1), 1)\\
&\quad -V_n(x_1, i_1, j \cdot \mathbf{1}(j \neq 1), x_2 + \mathbf{1}(j = 1), 0)]\\
&\quad +\mu_2[V_n(x_1, i_1, \mathbf{1}(j = 1), x_2 - 1, 1) - V_n(x_1, i_1, \mathbf{1}(j = 1), x_2, 0)]\} \leq 0,
\end{aligned}$$

because V_n satisfies (P1) (term multiplying ν_2) and (P4) (remaining terms). For property (P5) we have

$$\begin{aligned}
&W_{n+1}(x_1, i_1, 0, x_2, 1) - W_{n+1}(x_1, i_1, 2, x_2, 0)\\
&\quad = \beta\{\lambda[V_n(x_1 + 1, i_1, 0, x_2, 1) - V_n(x_1 + 1, i_1, 2, x_2, 0)]\\
&\quad +\nu_1[V_n(x_1, 0, 0, x_2 + i_1, 1) - V_n(x_1, 0, 2, x_2 + i_1, 0)]\\
&\quad +\nu_2[V_n(x_1, i_1, 0, x_2, 0) - V_n(x_1, i_1, 2, x_2, 0)]\\
&\quad +\mu_1[V_n(x_1, i_1, 0, x_2, 1) - V_n(x_1, i_1, 2, x_2, 0)]\\
&\quad +\mu_2[V_n(x_1, i_1, 0, x_2, 1) - V_n(x_1, i_1, 0, x_2, 0)]\}.
\end{aligned}$$

Because V_n satisfies (P2), the term multiplying μ_2 is nonnegative. Then, taking into account that $\mu_2 < \nu_2$, we get

$$\begin{aligned}
&W_{n+1}(x_1, i_1, 0, x_2, 1) - W_{n+1}(x_1, i_1, 2, x_2, 0)\\
&\quad \leq \beta\{\lambda[V_n(x_1 + 1, i_1, 0, x_2, 1) - V_n(x_1 + 1, i_1, 2, x_2, 0)]\\
&\quad +\nu_1[V_n(x_1, 0, 0, x_2 + i_1, 1) - V_n(x_1, 0, 2, x_2 + i_1, 0)]\\
&\quad +\nu_2[V_n(x_1, i_1, 0, x_2, 1) - V_n(x_1, i_1, 2, x_2, 0)]\\
&\quad +\mu_1[V_n(x_1, i_1, 0, x_2, 1) - V_n(x_1, i_1, 2, x_2, 0)]\} \leq 0,
\end{aligned}$$

because V_n satisfies (P5).

Proof of properties (Q1)-(Q8)

The proof is by induction on n and is structurally identical to the proof of (P1)-(P5). Assuming that W_n satisfies (Q1)-(Q8), we show that V_n and W_{n+1} satisfy these properties as well. According to the induction hypothesis, if there are jobs at Station 1 at time n and its dedicated server is available, the optimal policy assigns a job to that server, that is, for $x_1 \geq 1$ we have

$$V_n(x_1, 0, j, x_2) = V_n(x_1 - 1, 1, j, x_2). \tag{B.63}$$

Starting with the proof of (Q1) for V_n, let $(\tilde{\alpha}_1, \tilde{\alpha})$ be the optimal allocation for state $(x_1 + 1, i_1, j, x_2)$. Then, if $(\tilde{\alpha}_1, \tilde{\alpha})$ is feasible for state (x_1, i_1, j, x_2), we have from (2.4.24)

$$V_n(x_1, i_1, j, x_2) \leq W_n(x_1', \tilde{\alpha}_1, \tilde{\alpha}, x_2') \leq W_n(x_1' + 1, \tilde{\alpha}_1, \tilde{\alpha}, x_2') = V_n(x_1 + 1, i_1, j, x_2),$$

where $x_1' = x_1 - \tilde{\alpha}_1(1 - i_1) - \mathbf{1}(j = 0, \tilde{\alpha} = 1)$, $x_2' = x_2 - \mathbf{1}(j = 0, \tilde{\alpha} = 2)$, and the second inequality follows from property (Q1) for W_n. Next, we consider the cases for which $(\tilde{\alpha}_1, \tilde{\alpha})$ is not feasible for state (x_1, i_1, j, x_2).
Case 1. $x_1 = 0$, $i_1 = 0$. We have $\tilde{\alpha}_1 = 1$ and $\tilde{\alpha} = j$ if $j \neq 0$, $\tilde{\alpha} \neq 1$ if $j = 0$, so $(0, \tilde{\alpha})$ is feasible for state $(0, 0, j, x_2)$. Therefore,

$$V_n(0, 0, j, x_2) \leq W_n(0, 0, \tilde{\alpha}, x_2') \leq W_n(0, 1, \tilde{\alpha}, x_2') = V_n(1, 0, j, x_2),$$

where $x_2' = x_2 - \mathbf{1}(j = 0, \tilde{\alpha} = 2)$ and the second inequality follows from property (Q2) for W_n.
Case 2. $x_1 = 0$, $j = 0$, $i_1 = 1$, $\tilde{\alpha} = 1$. Then, because $(1, 0)$ is feasible for state $(0, 1, 0, x_2)$, we have

$$V_n(0, 1, 0, x_2) \leq W_n(0, 1, 0, x_2) \leq W_n(0, 1, 1, x_2) = V_n(1, 1, 0, x_2),$$

where the second inequality follows from property (Q3) for W_n.
Case 3. $x_1 = 1$, $i_1 = j = 0$, $\tilde{\alpha} = 1$. Because $\tilde{\alpha}_1 = 1$ and $(1, 0)$ is feasible for state $(1, 0, 0, x_2)$, we have

$$V_n(1, 0, 0, x_2) \leq W_n(0, 1, 0, x_2) \leq W_n(0, 1, 1, x_2) = V_n(2, 0, 0, x_2),$$

where the second inequality follows from property (Q3) for W_n.

Property (Q2) is a consequence of (Q1) and (B.63) (induction hypothesis) because

$$V_n(x_1, 0, j, x_2) \leq V_n(x_1 + 1, 0, j, x_2) = V_n(x_1, 1, j, x_2).$$

To prove (Q3), letting $\tilde{\alpha}_1$ be the optimal allocation of the dedicated server of Station 1 for state $(x_1, i_1, 1, x_2)$, we note that allocation $(\tilde{\alpha}_1, 0)$ is feasible for state $(x_1, i_1, 0, x_2)$. Therefore,

$$V_n(x_1, i_1, 0, x_2) \leq W_n(x_1', \tilde{\alpha}_1, 0, x_2) \leq W_n(x_1', \tilde{\alpha}_1, 1, x_2) = V_n(x_1, i_1, 1, x_2),$$

where $x_1' = x_1 - \tilde{\alpha}_1(1 - i_1)$ and the second inequality follows from property (Q3) for W_n.

The proofs of (Q4)-(Q6) are similar to those of (Q1)-(Q3), respectively. To see this, note that the only difference in the two sets of properties is that in the lefthand side of (Q4)-(Q6) there is one more job in Station 2 compared to (Q1)-(Q3). This difference has no effect on whether the optimal policy corresponding to the righthand side is feasible for the lefthand side, so the arguments in the proofs of (Q1)-(Q3) can be replicated for (Q4)-(Q6).

To prove property (Q4), we let $(\tilde{\alpha}_1, \tilde{\alpha})$ be the optimal allocation for state (x_1, i_1, j, x_2). Then, if $(\tilde{\alpha}_1, \tilde{\alpha})$ is feasible for state $(x_1 - 1, i_1, j, x_2 + 1)$, we have

$$V_n(x_1 - 1, i_1, j, x_2 + 1) \leq W_n(x_1' - 1, \tilde{\alpha}_1, \tilde{\alpha}, x_2' + 1) \leq W_n(x_1', \tilde{\alpha}_1, \tilde{\alpha}, x_2') = V_n(x_1, i_1, j, x_2),$$

where $x_1' = x_1 - \tilde{\alpha}_1(1 - i_1) - \mathbf{1}(j = 0, \tilde{\alpha} = 1)$, $x_2' = x_2 - \mathbf{1}(j = 0, \tilde{\alpha} = 2)$, and the second inequality follows from property (Q4) for W_n. Next, we consider the cases for which $(\tilde{\alpha}_1, \tilde{\alpha})$ is not feasible for state (x_1, i_1, j, x_2).
Case 1. $x_1 = 1$, $i_1 = 0$. We have $\tilde{\alpha}_1 = 1$ and $\tilde{\alpha} = j$ if $j \neq 0$, $\tilde{\alpha} \neq 1$ if $j = 0$, so $(0, \tilde{\alpha})$ is feasible for state $(0, 0, j, x_2 + 1)$. Therefore,

$$V_n(0, 0, j, x_2 + 1) \leq W_n(0, 0, \tilde{\alpha}, x_2' + 1) \leq W_n(0, 1, \tilde{\alpha}, x_2') = V_n(1, 0, j, x_2),$$

where $x_2' = x_2 - \mathbf{1}(j = 0, \tilde{\alpha} = 2)$ and the second inequality follows from property (Q5) for W_n.

Case 2. $x_1 = 1$, $j = 0$, $i_1 = 1$, $\tilde{\alpha} = 1$. Then, because $(1, 0)$ is feasible for state $(0, 1, 0, x_2 + 1)$, we have

$$V_n(0, 1, 0, x_2 + 1) \leq W_n(0, 1, 0, x_2 + 1) \leq W_n(0, 1, 1, x_2) = V_n(1, 1, 0, x_2),$$

where the second inequality follows from property (Q6) for W_n.

Case 3. $x_1 = 2$, $i_1 = j = 0$, $\tilde{\alpha} = 1$. Because $\tilde{\alpha}_1 = 1$ and $(1, 0)$ is feasible for state $(1, 0, 0, x_2 + 1)$, we have

$$V_n(1, 0, 0, x_2 + 1) \leq W_n(0, 1, 0, x_2 + 1) \leq W_n(0, 1, 1, x_2) = V_n(2, 0, 0, x_2),$$

where the second inequality follows from property (Q6) for W_n.

Property (Q5) follows from (Q4) and (B.63) because

$$V_n(x_1, 0, j, x_2 + 1) \leq V_n(x_1 + 1, 0, j, x_2) = V_n(x_1, 1, j, x_2).$$

To prove (Q6), letting $\tilde{\alpha}_1$ be the optimal allocation of the dedicated server of Station 1 for state $(x_1, i_1, 1, x_2)$, we note that allocation $(\tilde{\alpha}_1, 0)$ is feasible for state $(x_1, i_1, 0, x_2 + 1)$. Therefore,

$$V_n(x_1, i_1, 0, x_2 + 1) \leq W_n(x_1', \tilde{\alpha}_1, 0, x_2 + 1) \leq W_n(x_1', \tilde{\alpha}_1, 1, x_2) = V_n(x_1, i_1, 1, x_2),$$

where $x_1' = x_1 - \tilde{\alpha}_1(1 - i_1)$ and the second inequality follows from property (Q6) for W_n.

As a consequence of the induction hypothesis, property (Q7) for V_n is satisfied with equality. We also use the induction hypothesis to prove (Q8) for $x_1 \geq 1$, in which case we get

$$V_n(x_1, 1, 0, x_2) \leq V_n(x_1 - 1, 1, 1, x_2) = V_n(x_1, 0, 1, x_2),$$

where the inequality is due to the fact that assigning the flexible server to station 1 may not be optimal. For $x_2 = 0$, noting that allocation $(1, 0)$ is feasible for state $(0, 1, 0, x_2)$ we obtain

$$V_n(0, 1, 0, x_2) \leq W_n(0, 1, 0, x_2) \leq W_n(0, 0, 1, x_2) = V_n(0, 0, 1, x_2),$$

where the second inequality follows from property (Q8) for W_n.

Finally, we show the same properties for the function W_{n+1} satisfies properties (Q1)-(Q8) by using (2.4.25) and (2.4.26). For property (Q1) we have

$$\begin{aligned}
&W_{n+1}(x_1, i_1, j, x_2) - W_{n+1}(x_1 + 1, i_1, j, x_2) = -h_1 \\
&\quad +\beta\{\lambda[V_n(x_1 + 1, i_1, j, x_2) - V_n(x_1 + 2, i_1, j, x_2)] \\
&\quad +\nu_1[V_n(x_1, 0, j, x_2 + i_1) - V_n(x_1 + 1, 0, j, x_2 + i_1)] \\
&\quad +\nu_2[V_n(x_1, i_1, j, (x_2 - 1)^+) - V_n(x_1 + 1, i_1, j, (x_2 - 1)^+)] \\
&\quad +\mu_1[V_n(x_1, i_1, j \cdot \mathbf{1}(j \neq 1), x_2 + \mathbf{1}(j = 1)) \\
&\quad -V_n(x_1 + 1, i_1, j \cdot \mathbf{1}(j \neq 1), x_2 + \mathbf{1}(j = 1))] \\
&\quad +\mu_2[V_n(x_1, i_1, \mathbf{1}(j = 1), x_2) - V_n(x_1 + 1, i_1, \mathbf{1}(j = 1), x_2)]\} < 0,
\end{aligned}$$

because V_n satisfies (Q1). For property (Q2) we have

$$\begin{aligned}
&W_{n+1}(x_1,0,j,x_2) - W_{n+1}(x_1,1,j,x_2) = -h_1 \\
&\quad +\beta\{\lambda[V_n(x_1+1,0,j,x_2) - V_n(x_1+1,1,j,x_2)] \\
&\quad +\nu_1[V_n(x_1,0,j,x_2) - V_n(x_1,0,j,x_2+1)] \\
&\quad +\nu_2[V_n(x_1,0,j,(x_2-1)^+) - V_n(x_1,1,j,(x_2-1)^+)] \\
&\quad +\mu_1[V_n(x_1,0,j\cdot\mathbf{1}(j\neq 1),x_2+\mathbf{1}(j=1)) - V_n(x_1,1,j\cdot\mathbf{1}(j\neq 1),x_2+\mathbf{1}(j=1))] \\
&\quad +\mu_2[V_n(x_1,0,\mathbf{1}(j=1),x_2) - V_n(x_1,1,\mathbf{1}(j=1),x_2)]\} < 0,
\end{aligned}$$

because of (2.4.28) (term multiplying ν_1) and the fact thatV_n satisfies (Q2) (remaining terms). For property (Q3) we have

$$\begin{aligned}
&W_{n+1}(x_1,i_1,0,x_2) - W_{n+1}(x_1,i_1,1,x_2) = -h_1 \\
&\quad +\beta\{\lambda[V_n(x_1+1,i_1,0,x_2) - V_n(x_1+1,i_1,1,x_2)] \\
&\quad +\nu_1[V_n(x_1,0,0,x_2+i_1) - V_n(x_1,0,1,x_2+i_1)] \\
&\quad +\nu_2[V_n(x_1,i_1,0,(x_2-1)^+) - V_n(x_1,i_1,1,(x_2-1)^+)] \\
&\quad +\mu_1[V_n(x_1,i_1,0,x_2) - V_n(x_1,i_1,0,x_2+1)] \\
&\quad +\mu_2[V_n(x_1,i_1,0,x_2) - V_n(x_1,i_1,1,x_2)]\} < 0,
\end{aligned}$$

because of (2.4.28) (term multiplying μ_1) and the fact that V_n satisfies (Q3) (remaining terms). For property (Q4) we have

$$\begin{aligned}
&W_{n+1}(x_1-1,i_1,j,x_2+1) - W_{n+1}(x_1,i_1,j,x_2) = h_2 - h_1 \\
&\quad +\beta\{\lambda[V_n(x_1,i_1,j,x_2+1) - V_n(x_1+1,i_1,j,x_2)] \\
&\quad +\nu_1[V_n(x_1-1,0,j,x_2+1+i_1) - V_n(x_1,0,j,x_2+i_1)] \\
&\quad +\nu_2[V_n(x_1-1,i_1,j,x_2) - V_n(x_1,i_1,j,(x_2-1)^+)] \\
&\quad +\mu_1[V_n(x_1-1,i_1,j\cdot\mathbf{1}(j\neq 1),x_2+1+\mathbf{1}(j=1)) \\
&\quad -V_n(x_1,i_1,j\cdot\mathbf{1}(j\neq 1),x_2+\mathbf{1}(j=1))] \\
&\quad +\mu_2[V_n(x_1-1,i_1,\mathbf{1}(j=1),x_2+1) - V_n(x_1,i_1,\mathbf{1}(j=1),x_2)]\} \leq 0,
\end{aligned}$$

because V_n satisfies (Q4) and (Q1) (term multiplying ν_2 when $x_2 = 0$). For property (Q5) we have

$$\begin{aligned}
&W_{n+1}(x_1,0,j,x_2+1) - W_{n+1}(x_1,1,j,x_2) = h_2 - h_1 \\
&\quad +\beta\{\lambda[V_n(x_1+1,0,j,x_2+1) - V_n(x_1+1,1,j,x_2)] \\
&\quad +\nu_2[V_n(x_1,0,j,x_2) - V_n(x_1,1,j,(x_2-1)^+)] \\
&\quad +\mu_1[V_n(x_1,0,j\cdot\mathbf{1}(j\neq 1),x_2+1+\mathbf{1}(j=1)) \\
&\quad -V_n(x_1,1,j\cdot\mathbf{1}(j\neq 1),x_2+\mathbf{1}(j=1))] \\
&\quad +\mu_2[V_n(x_1,0,\mathbf{1}(j=1),x_2+1) - V_n(x_1,1,\mathbf{1}(j=1),x_2)]\} \leq 0,
\end{aligned}$$

because V_n satisfies (Q5) and (Q2) (term multiplying ν_2 when $x_2 = 0$). For property (Q6) we have

$$\begin{aligned}W_{n+1}(x_1,i_1,0,x_2+1)-W_{n+1}(x_1,i_1,1,x_2)&=h_2-h_1\\ &+\beta\{\lambda[V_n(x_1+1,i_1,0,x_2+1)-V_n(x_1+1,i_1,1,x_2)]\\ &+\nu_1[V_n(x_1,0,0,x_2+1+i_1)-V_n(x_1,0,1,x_2+i_1)]\\ &+\nu_2[V_n(x_1,i_1,0,x_2)-V_n(x_1,i_1,1,(x_2-1)^+)]\\ &+\mu_2[V_n(x_1,i_1,0,x_2+1)-V_n(x_1,i_1,1,x_2)]\}\le 0,\end{aligned}$$

because V_n satisfies (Q6) and (Q3) (term multiplying ν_2 when $x_2=0$). For property (Q7) we have

$$\begin{aligned}&W_{n+1}(x_1-1,1,j,x_2)-W_{n+1}(x_1,0,j,x_2)\\ &\quad=\beta\{\lambda[V_n(x_1,1,j,x_2)-V_n(x_1+1,0,j,x_2)]\\ &\quad+\nu_1[V_n(x_1-1,0,j,x_2+1)-V_n(x_1,0,j,x_2)]\\ &\quad+\nu_2[V_n(x_1-1,1,j,(x_2-1)^+)-V_n(x_1,0,j,(x_2-1)^+)]\\ &\quad+\mu_1[V_n(x_1-1,1,j\cdot\mathbf{1}(j\neq 1),x_2+\mathbf{1}(j=1))\\ &\quad-V_n(x_1,0,j\cdot\mathbf{1}(j\neq 1),x_2+\mathbf{1}(j=1))]\\ &\quad+\mu_2[V_n(x_1-1,1,\mathbf{1}(j=1),x_2)-V_n(x_1,0,\mathbf{1}(j=1),x_2)]\}\le 0,\end{aligned}$$

because V_n satisfies (Q4) (term multiplying ν_1) and (Q7) (remaining terms). For property (Q8) we have

$$\begin{aligned}&W_{n+1}(x_1,1,0,x_2)-W_{n+1}(x_1,0,1,x_2)\\ &\quad=\beta\{\lambda[V_n(x_1+1,1,0,x_2)-V_n(x_1+1,0,1,x_2)]\\ &\quad+\nu_1[V_n(x_1,0,0,x_2+1)-V_n(x_1,0,1,x_2)]\\ &\quad+\nu_2[V_n(x_1,1,0,(x_2-1)^+)-V_n(x_1,0,1,(x_2-1)^+)]\\ &\quad+\mu_1[V_n(x_1,1,0,x_2)-V_n(x_1,0,0,x_2+1)]\\ &\quad+\mu_2[V_n(x_1,1,0,x_2)-V_n(x_1,0,1,x_2)]\}.\end{aligned}$$

Because V_n satisfies (Q5), the term multiplying μ_1 is nonnegative. Then, taking into account that $\mu_1<\nu_1$, we get

$$\begin{aligned}&W_{n+1}(x_1,1,0,x_2)-W_{n+1}(x_1,0,1,x_2)\\ &\quad\le\beta\{\lambda[V_n(x_1+1,1,0,x_2)-V_n(x_1+1,0,1,x_2)]\\ &\quad+\nu_1[V_n(x_1,1,0,x_2)-V_n(x_1,0,1,x_2)]\\ &\quad+\nu_2[V_n(x_1,1,0,(x_2-1)^+)-V_n(x_1,0,1,(x_2-1)^+)]\\ &\quad+\mu_2[V_n(x_1,1,0,x_2)-V_n(x_1,0,1,x_2)]\},\end{aligned}$$

because V_n satisfies (Q8).

Appendix C

Proof of Proposition 3.1

Differentiating (3.4.4)-(3.4.9) with respect to Q and K we get the following expressions for the second-order derivatives of the profit function. For $Q + K \leq I$,

$$\frac{\partial^2 \Pi}{\partial Q^2} = -(r+p-h)E\left[U^2 f(QU+K)\right], \tag{C.1}$$

$$\frac{\partial^2 \Pi}{\partial K^2} = -(r+p-h)E\left[f(QU+K)\right], \tag{C.2}$$

$$\frac{\partial^2 \Pi}{\partial Q \partial K} = -(r+p-h)E\left[Uf(QU+K)\right], \tag{C.3}$$

for $I - K < Q < I$,

$$\frac{\partial^2 \Pi}{\partial Q^2} = -(r+p-h)\int_0^{\frac{I-K}{Q}} u^2 f(Qu+K)g(u)du, \tag{C.4}$$

$$\frac{\partial^2 \Pi}{\partial K^2} = -(r+p-h)\int_0^{\frac{I-K}{Q}} f(Qu+K)g(u)du, \tag{C.5}$$

$$\frac{\partial^2 \Pi}{\partial Q \partial K} = -(r+p-h)\int_0^{\frac{I-K}{Q}} uf(Qu+K)g(u)du, \tag{C.6}$$

and for $Q \geq I$,

$$\frac{\partial^2 \Pi}{\partial Q^2} = -(r+p-h)\left[\int_0^{\frac{I-K}{Q}} u^2 f(Qu+K)g(u)du + \int_{\frac{I}{Q}}^{1} u^2 f(Qu)g(u)du\right], \tag{C.7}$$

$$\frac{\partial^2 \Pi}{\partial K^2} = -(r+p-h)\int_0^{\frac{I-K}{Q}} f(Qu+K)g(u)du, \tag{C.8}$$

$$\frac{\partial^2 \Pi}{\partial Q \partial K} = -(r+p-h)\int_0^{\frac{I-K}{Q}} uf(Qu+K)g(u)du. \tag{C.9}$$

Let H be the Hessian matrix of the profit function, that is,

$$H = \begin{bmatrix} \frac{\partial^2 \Pi}{\partial Q^2} & \frac{\partial^2 \Pi}{\partial Q \partial K} \\ \frac{\partial \Pi^2}{\partial K \partial Q} & \frac{\partial^2 \Pi}{\partial K^2} \end{bmatrix}.$$

Then, using (C.1)-(C.9) we get the following expressions for its determinant. For $Q+K \leq I$,

$$\begin{aligned}\det(H) &= (r+p-h)^2 E\left[U^2 f\left(QU+K\right)\right] E\left[f\left(QU+K\right)\right] \\ &\quad -(r+p-h)^2 \left[E\left[U f\left(QU+K\right)\right]\right]^2,\end{aligned} \tag{C.10}$$

for $I-K < Q < I$,

$$\begin{aligned}\det(H) &= (r+p-h)^2 \left[\int_0^{\frac{I-K}{Q}} u^2 f\left(Qu+K\right) g(u) du\right] \left[\int_0^{\frac{I-K}{Q}} f\left(Qu+K\right) g(u) du\right] \\ &\quad -(r+p-h)^2 \left[\int_0^{\frac{I-K}{Q}} u f\left(Qu+K\right) g(u) du\right]^2,\end{aligned} \tag{C.11}$$

and for $Q \geq I$,

$$\begin{aligned}\det(H) &= (r+p-h)^2 \left[\int_0^{\frac{I-K}{Q}} u^2 f\left(Qu+K\right) g(u) du\right] \left[\int_0^{\frac{I-K}{Q}} f\left(Qu+K\right) g(u) du\right] \\ &\quad +(r+p-h)^2 \left[\int_0^{\frac{I-K}{Q}} f\left(Qu+K\right) g(u) du\right] \left[\int_{\frac{I}{Q}}^{1} u^2 f\left(Qu\right) g(u) du\right] \\ &\quad -(r+p-h)^2 \left[\int_0^{\frac{I-K}{Q}} u f\left(Qu+K\right) g(u) du\right]^2.\end{aligned} \tag{C.12}$$

Applying the Cauchy-Schwarz inequality, given by

$$\left[\int_a^b \psi_1(u)\psi_2(u) du\right]^2 \leq \left[\int_a^b \left[\psi_1(u)\right]^2 du\right] \left[\int_a^b \left[\psi_2(u)\right]^2 du\right],$$

with $a = 0$, $b = \min\{(I-K)/Q, 1\}$, $\psi_1(u) = u\sqrt{f(Qu+K)}$, and $\psi_2(u) = \sqrt{f(Qu+K)}$, we get from (C.10)-(C.12) that $\det(H) \geq 0$, which combined with $\partial^2\Pi/\partial Q^2 \leq 0$ and $\partial^2\Pi/\partial K^2 \leq 0$ proves that the profit function is concave.

Proof of Proposition 3.3

To prove that the expected profit is concave it suffices to show that the profit function is concave for every realization of U_1, U_2, X. For realizations u_1, u_2, x of the aforementioned random variables we define sets $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3$ as follows.

$$\begin{aligned}\mathcal{R}_1 &= \{(Q_1, Q_2, K): \; x \leq Q_1u_1 + Q_2u_2\}, \\ \mathcal{R}_2 &= \{(Q_1, Q_2, K): \; Q_1u_1 + Q_2u_2 < x \leq Q_1u_1 + Q_2u_2 + K\}, \\ \mathcal{R}_3 &= \{(Q_1, Q_2, K): \; x > Q_1u_1 + Q_2u_2 + K\}.\end{aligned}$$

Then, the profit resulting from ordering Q_1, Q_2 from the primary suppliers and reserving K with the backup supplier, denoted by $\tilde{\Pi}(Q_1, Q_2, K)$, is equal to

$$\tilde{\Pi}(Q_1, Q_2, K) = -c_1 Q_1 u_1 - c_2 Q_2 u_2 - c_r K + \tilde{L}(Q_1, Q_2, K),$$

where

$$\begin{aligned} \tilde{L}(Q_1, Q_2, K) &= rx + h(Q_1 u_1 + Q_2 u_2 - x), \text{ if } (Q_1, Q_2, K) \in \mathcal{R}_1, && \text{(C.13)} \\ &= rx - c_e(x - Q_1 u_1 - Q_2 u_2), \text{ if } (Q_1, Q_2, K) \in \mathcal{R}_2, && \text{(C.14)} \\ &= -c_e K + r(Q_1 u_1 + Q_2 u_2 + K) - p(x - Q_1 u_1 - Q_2 u_2 - K), \\ &\qquad \text{if } (Q_1, Q_2, K) \in \mathcal{R}_3. && \text{(C.15)} \end{aligned}$$

To prove the concavity of function $\tilde{\Pi}$ we will show that for any vectors $T_i = (Q_{1i}, Q_{2i}, K_i)$, $i = 1, 2$, and any $0 < \lambda < 1$ we have

$$\tilde{\Pi}(\lambda T_1 + (1-\lambda) T_2) \geq \lambda \tilde{\Pi}(T_1) + (1-\lambda)\tilde{\Pi}(T_2). \tag{C.16}$$

This is trivial when $T_1, T_2 \in \mathcal{R}_i$ for some $i = 1, 2, 3$, because $\lambda T_1 + (1-\lambda)T_2 \in \mathcal{R}_i$ as well and (C.16) holds with equality. For cases with T_1, T_2 belonging to different sets, let LR denote the difference between the left-hand side and the right-hand side of (C.16).

<u>Case 1:</u> $T_1 \in \mathcal{R}_1$, $T_2 \in \mathcal{R}_2$

For λ such that $\lambda T_1 + (1-\lambda)T_2 \in \mathcal{R}_1$ we get

$$LR = (1-\lambda)(c_e - h)(x - Q_{12}u_1 - Q_{22}u_2) > 0,$$

because $c_e > h$ and $T_2 \in \mathcal{R}_2$. For λ such that $\lambda T_1 + (1-\lambda)T_2 \in \mathcal{R}_2$ we get

$$LR = \lambda(c_e - h)(Q_{11}u_1 + Q_{21}u_2 - x) \geq 0,$$

because $c_e > h$ and $T_1 \in \mathcal{R}_1$.

<u>Case 2:</u> $T_1 \in \mathcal{R}_2$, $T_2 \in \mathcal{R}_3$

For λ such that $\lambda T_1 + (1-\lambda)T_2 \in \mathcal{R}_2$ we get

$$LR = (1-\lambda)(p + r - c_e)(x - Q_{12}u_1 - Q_{22}u_2 - K_2) > 0,$$

because $p + r > c_e$ and $T_2 \in \mathcal{R}_3$. For λ such that $\lambda T_1 + (1-\lambda)T_2 \in \mathcal{R}_3$ we get

$$LR = \lambda(p + r - c_e)(Q_{11}u_1 + Q_{21}u_2 + K_1 - x) \geq 0,$$

because $p + r > c_e$ and $T_1 \in \mathcal{R}_2$.

<u>Case 3:</u> $T_1 \in \mathcal{R}_1$, $T_2 \in \mathcal{R}_3$

For λ such that $\lambda T_1 + (1-\lambda)T_2 \in \mathcal{R}_1$ we get

$$LR = (1-\lambda)\left[(p + r - h)(x - Q_{12}u_1 - Q_{22}u_2 - K_2) + (c_e - h)K_2\right] > 0,$$

because $p + r > h$, $c_e > h$, and $T_2 \in \mathcal{R}_3$. For λ such that $\lambda T_1 + (1-\lambda)T_2 \in \mathcal{R}_2$ we get

$$LR = (1-\lambda)(p + r - c_e)(x - Q_{12}u_1 - Q_{22}u_2 - K_2) + \lambda(c_e - h)(Q_{11}u_1 + Q_{21}u_2 - x) > 0,$$

because $p + r > c_e$, $c_e > h$, $T_2 \in \mathcal{R}_3$, and $T_1 \in \mathcal{R}_1$. For λ such that $\lambda T_1 + (1-\lambda)T_2 \in \mathcal{R}_3$ we get

$$LR = \lambda\left[(p + r - h)(Q_{11}u_1 - Q_{21}u_2 - x) + (p + r - c_e)K_1\right] \geq 0,$$

because $p + r > h$, $p + r > c_e$, and $T_1 \in \mathcal{R}_1$.

Proof of Proposition 3.4

We will show that the expected profit is concave for every realization of U_1, X_1, U_2, X_2. For realizations u_1, x_1, u_2, x_2 of the aforementioned random variables we denote by $\tilde{\Pi}(Q_1, Q_2, K)$ the profit resulting from ordering Q_1, Q_2 from the primary suppliers and reserving K from the backup supplier. Then,

$$\tilde{\Pi}(Q_1, Q_2, K) = -c_1 Q_1 u_1 - c_2 Q_2 u_2 - c_r K + \tilde{\Pi}_1(Q_1, K) + \tilde{\Pi}_2(Q_1, Q_2, K),$$

where $\tilde{\Pi}_1(Q_1, K)$ and $\tilde{\Pi}_2(Q_1, Q_2, K)$ are the realized profits from product 1 and 2, respectively. To obtain expressions for these profits we define sets $\mathcal{R}_{ij}$ and functions $\tilde{L}_i$, $i = 1, 2$, $j = 1, 2, 3$, as follows.

$$\begin{aligned}
\mathcal{R}_{i1} &= \{(Q, K) : \ x_i \leq Q u_i\}, \\
\mathcal{R}_{i2} &= \{(Q, K) : \ Q u_i < x_i \leq Q u_i + K\}, \\
\mathcal{R}_{i3} &= \{(Q, K) : \ x_i > Q u_i + K\},
\end{aligned}$$

and

$$\begin{aligned}
\tilde{L}_i(Q, K) &= r_i x_i + h_i(Q u_i - x_i), \ \text{if } (Q, K) \in \mathcal{R}_{i1}, && \text{(C.17)} \\
&= r_i x_i - c_{ei}(x_i - Q u_i), \ \text{if } (Q, K) \in \mathcal{R}_{i2}, && \text{(C.18)} \\
&= -c_{ei} K + r_i(Q u_i + K) - p_i(x_i - Q u_i - K), \ \text{if } (Q, K) \in \mathcal{R}_{i3}. && \text{(C.19)}
\end{aligned}$$

Then,

$$\begin{aligned}
\tilde{\Pi}_1(Q_1, K) &= \tilde{L}_1(Q_1, K), && \text{(C.20)} \\
\tilde{\Pi}_2(Q_1, Q_2, K) &= \tilde{L}_2(Q_2, K), \ \text{if } (Q_1, K) \in \mathcal{R}_{11}, && \text{(C.21)} \\
&= \tilde{L}_2(Q_2, K - x_1 + Q_1 u_1), \ \text{if } (Q_1, K) \in \mathcal{R}_{12}, && \text{(C.22)} \\
&= \tilde{L}_2(Q_2, 0), \ \text{if } (Q_1, K) \in \mathcal{R}_{13}. && \text{(C.23)}
\end{aligned}$$

To prove the concavity of function $\tilde{\Pi}$ we will show that for any vectors $T_i = (Q_{1i}, Q_{2i}, K_i)$, $i = 1, 2$, and any $0 < \lambda < 1$ we have $LR_1 + LR_2 \geq 0$, where

$$\begin{aligned}
LR_1 &= \tilde{\Pi}_1(\lambda \tilde{T}_1 + (1-\lambda)\tilde{T}_2) - \lambda \tilde{\Pi}_1(\tilde{T}_1) - (1-\lambda)\tilde{\Pi}_1(\tilde{T}_2), && \text{(C.24)} \\
LR_2 &= \tilde{\Pi}_2(\lambda T_1 + (1-\lambda) T_2) - \lambda \tilde{\Pi}_2(T_1) - (1-\lambda)\tilde{\Pi}_2(T_2), && \text{(C.25)}
\end{aligned}$$

where $\tilde{T}_i = (Q_{1i}, K_i)$. Because functions $\tilde{L}_1, \tilde{L}_2$ are special cases of function $\tilde{L}$ defined in the proof of Proposition 3.3 for two suppliers and one product, they are jointly concave in Q, K. Therefore, we get from (C.20) and (C.24) that $LR_1 \geq 0$. As for LR_2 we see from (C.21)-(C.23) and (C.25) that it is given by the following general expression.

$$LR_2 = \tilde{L}_2(\lambda Q_{21} + (1-\lambda) Q_{22}, \tilde{K}) - \lambda \tilde{L}_2(Q_{21}, \tilde{K}_1) - (1-\lambda)\tilde{L}_2(Q_{22}, \tilde{K}_2). \qquad \text{(C.26)}$$

Taking also into account that $\tilde{L}_2(Q, K)$ is nondecreasing in K, which can be easily derived from (C.17)-(C.19) and the definition of $\mathcal{R}_{2j}$, $j = 1, 2, 3$, we have $LR_2 \geq 0$ whenever $\tilde{K} \geq$

$\lambda\tilde{K}_1 + (1-\lambda)\tilde{K}_2$. This condition is satisfied with equality when $\tilde{T}_1, \tilde{T}_2 \in \mathcal{R}_{1j}$ for some $j = 1, 2, 3$, which implies that $\lambda\tilde{T}_1 + (1-\lambda)\tilde{T}_2 \in \mathcal{R}_{1j}$ as well. It is also easy to verify that the condition is satisfied in the following cases: i) $\tilde{T}_1 \in \mathcal{R}_{11}$, $\tilde{T}_2 \in \mathcal{R}_{12}$, and ii) $\tilde{T}_1 \in \mathcal{R}_{11}$, $\tilde{T}_2 \in \mathcal{R}_{13}$, and $x_1 \leq (\lambda Q_{11} + (1-\lambda)Q_{12})\, u_1 + (1-\lambda)K_2$, that is, $\lambda\tilde{T}_1 + (1-\lambda)\tilde{T}_2$ belongs to $\mathcal{R}_{11}$ or to the subset of $\mathcal{R}_{12}$ specified by the last inequality. For all other cases we have $\tilde{K} \leq \lambda\tilde{K}_1 + (1-\lambda)\tilde{K}_2$, so we need to consider cases with $\tilde{T}_1 \in \mathcal{R}_{11}$, $\tilde{T}_2 \in \mathcal{R}_{13}$ and $\tilde{T}_1 \in \mathcal{R}_{12}$, $\tilde{T}_2 \in \mathcal{R}_{13}$, both implying that $\tilde{K}_2 = 0$, and all possible combinations of sets $\mathcal{R}_{2j}$, $j = 1, 2, 3$, to which $(Q_{21}, \tilde{K}_1)$, $(Q_{22}, \tilde{K}_2)$, and $(\lambda Q_{21} + (1-\lambda)Q_{22}, \tilde{K})$ may belong. In case $(\lambda Q_{21} + (1-\lambda)Q_{22}, \tilde{K}) \in \mathcal{R}_{21}$ or $\mathcal{R}_{22}$, we also have $(\lambda Q_{21} + (1-\lambda)Q_{22}, \lambda\tilde{K}_1 + (1-\lambda)\tilde{K}_2) \in \mathcal{R}_{21}$ or $\mathcal{R}_{22}$, respectively, because $\tilde{K} \leq \lambda\tilde{K}_1 + (1-\lambda)\tilde{K}_2$. Then, we see from (C.17) and (C.18) that $\tilde{L}_2(\lambda Q_{21} + (1-\lambda)Q_{22}, \tilde{K}) = \tilde{L}_2(\lambda Q_{21} + (1-\lambda)Q_{22}, \lambda\tilde{K}_1 + (1-\lambda)\tilde{K}_2)$, and $LR_2 \geq 0$ follows from the concavity of $\tilde{L}_2$. Therefore, we need to compute LR_2 for cases with $(\lambda Q_{21} + (1-\lambda)Q_{22}, \tilde{K}) \in \mathcal{R}_{23}$ and $\tilde{K}_2 = 0$, which means that $(Q_{22}, \tilde{K}_2) \notin \mathcal{R}_{22}$. Note that $(Q_{21}, \tilde{K}_1)$ and $(Q_{22}, \tilde{K}_2)$ cannot both belong to $\mathcal{R}_{21}$ because we would have $x_2 \leq (\lambda Q_{21} + (1-\lambda)Q_{22})u_2$, contradicting the fact that $(\lambda Q_{21} + (1-\lambda)Q_{22}, \tilde{K}) \in \mathcal{R}_{23}$. Then, using (C.26), (C.17)-(C.19), and the definitions of $\mathcal{R}_{2j}$, $j = 1, 2, 3$, we get the following for the remaining cases. For $(Q_{21}, \tilde{K}_1) \in \mathcal{R}_{21}$, $(Q_{22}, \tilde{K}_2) \in \mathcal{R}_{23}$,

$$LR_2 = (p_2 + r_2 - c_{e2})\tilde{K} + \lambda(p_2 + r_2 - h_2)(Q_{21}u_2 - x_2) \geq 0,$$

for $(Q_{21}, \tilde{K}_1) \in \mathcal{R}_{22}$, $(Q_{22}, \tilde{K}_2) \in \mathcal{R}_{21}$,

$$\begin{aligned} LR_2 &= (p_2 + r_2 - c_{e2})\left[\tilde{K} + \lambda(Q_{21}u_2 - x_2)\right] + (1-\lambda)(p_2 + r_2 - h_2)(Q_{22}u_2 - x_2) \\ &\geq (p_2 + r_2 - c_{e2})(\tilde{K} - \lambda\tilde{K}_1), \end{aligned} \tag{C.27}$$

for $(Q_{21}, \tilde{K}_1) \in \mathcal{R}_{22}$, $(Q_{22}, \tilde{K}_2) \in \mathcal{R}_{23}$,

$$LR_2 = (p_2 + r_2 - c_{e2})\left[\tilde{K} + \lambda(Q_{21}u_2 - x_2)\right] \geq (p_2 + r_2 - c_{e2})(\tilde{K} - \lambda\tilde{K}_1), \tag{C.28}$$

for $(Q_{21}, \tilde{K}_1) \in \mathcal{R}_{23}$, $(Q_{22}, \tilde{K}_2) \in \mathcal{R}_{21}$,

$$\begin{aligned} LR_2 &= (p_2 + r_2 - c_{e2})(\tilde{K} - \lambda\tilde{K}_1) + (1-\lambda)(p_2 + r_2 - h_2)(Q_{22}u_2 - x_2) \\ &\geq (p_2 + r_2 - c_{e2})(\tilde{K} - \lambda\tilde{K}_1), \end{aligned} \tag{C.29}$$

and for $(Q_{21}, \tilde{K}_1) \in \mathcal{R}_{23}$, $(Q_{22}, \tilde{K}_2) \in \mathcal{R}_{23}$,

$$LR_2 = (p_2 + r_2 - c_{e2})(\tilde{K} - \lambda\tilde{K}_1). \tag{C.30}$$

To deal with the last four cases (Eqs (C.27)-(C.30)) we need to get expressions for LR_1 as well.

<u>Case 1.</u> $\tilde{T}_1 \in \mathcal{R}_{12}$, $\tilde{T}_2 \in \mathcal{R}_{13}$

For $\lambda\tilde{T}_1 + (1-\lambda)\tilde{T}_2 \in \mathcal{R}_{12}$ we have $\tilde{K} - \lambda\tilde{K}_1 = (1-\lambda)(K_2 - x_1 + Q_{12}u_1)$ and by replicating the arguments used in the proof of Proposition 3.3 (Case 2)

$$LR_1 = (1-\lambda)(p_1 + r_1 - c_{e1})(x_1 - Q_{12}u_1 - K_2),$$

which combined with each one of (C.27)-(C.30) yield

$$LR_1 + LR_2 \geq (1-\lambda)\left[(p_1 + r_1 - c_{e1}) - (p_2 + r_2 - c_{e2})\right](x_1 - Q_{12}u_1 - K_2) \geq 0,$$

because $\tilde{T}_2 \in \mathcal{R}_{13}$. For $\lambda\tilde{T}_1 + (1-\lambda)\tilde{T}_2 \in \mathcal{R}_{13}$ we have $\tilde{K} - \lambda\tilde{K}_1 = \lambda(x_1 - K_1 - Q_{11}u_1)$ and

$$LR_1 = \lambda(p_1 + r_1 - c_{e1})(K_1 - x_1 + Q_{11}u_1),$$

so we get

$$LR_1 + LR_2 \geq \lambda\left[(p_1 + r_1 - c_{e1}) - (p_2 + r_2 - c_{e2})\right](K_1 - x_1 + Q_{11}u_1) \geq 0,$$

because $\tilde{T}_1 \in \mathcal{R}_{12}$.

Case 2: $\tilde{T}_1 \in \mathcal{R}_{11}$, $\tilde{T}_2 \in \mathcal{R}_{13}$

For $\lambda\tilde{T}_1 + (1-\lambda)\tilde{T}_2 \in \mathcal{R}_{12}$ we have $\tilde{K} - \lambda\tilde{K}_1 = (1-\lambda)(K_2 - x_1 + Q_{12}u_1) + \lambda(Q_{11}u_1 - x_1)$ and by replicating the arguments used in the proof of Proposition 3.3 (Case 3)

$$LR_1 = (1-\lambda)(p_1 + r_1 - c_{e1})(x_1 - Q_{12}u_1 - K_2) + \lambda(c_{e1} - h_1)(Q_{11}u_1 - x_1),$$

which combined with each one of (C.27)-(C.30) yield

$$\begin{aligned} LR_1 + LR_2 \ \geq\ & (1-\lambda)\left[(p_1 + r_1 - c_{e1}) - (p_2 + r_2 - c_{e2})\right](x_1 - Q_{12}u_1 - K_2) \\ & + \lambda(p_2 + r_2 - c_{e2} + c_{e1} - h_1)(Q_{11}u_1 - x_1) \geq 0, \end{aligned}$$

because $\tilde{T}_2 \in \mathcal{R}_{13}$ and $\tilde{T}_1 \in \mathcal{R}_{11}$. For $\lambda\tilde{T}_1 + (1-\lambda)\tilde{T}_2 \in \mathcal{R}_{13}$ we have $\tilde{K} - \lambda\tilde{K}_1 = -\lambda K_1$ and

$$LR_1 = \lambda(p_1 + r_1 - c_{e1})K_1 + \lambda(p_1 + r_1 - h_1)(Q_{11}u_1 - x_1),$$

so we get

$$\begin{aligned} LR_1 + LR_2 \ =\ & \lambda\left[(p_1 + r_1 - c_{e1}) - (p_2 + r_2 - c_{e2})\right]K_1 \\ & + \lambda(p_1 + r_1 - h_1)(Q_{11}u_1 - x_1) \geq 0, \end{aligned}$$

because $\tilde{T}_1 \in \mathcal{R}_{11}$.

Bibliography

[1] A.K. Agrawala, E.G. Coffman, M.R. Garey, and S.K. Tripathi, "A stochastic optimization algorithm minimizing expected flow times on uniform processors," *IEEE Transactions on Computers*, vol. 33, no. 4, pp. 351-356, 1984.

[2] N. Agrawal and S. Nahmias, "Rationalization of the supplier base in the presence of yield uncertainty," *Production and Operations Management*, vol. 6, no. 3, pp. 291-308, 1997.

[3] H.-S. Ahn, I. Duenyas, and R. Zhang, "Optimal stochastic scheduling of a two-stage tandem queue with parallel servers," *Advances in Applied Probability*, vol. 31, no. 4, pp. 1095-1117, 1999.

[4] H.-S. Ahn, I. Duenyas, and M.E. Lewis, "The optimal control of a two-stage tandem queueing system with flexible servers," *Probability in the Engineering and Informational Sciences*, vol. 16, no. 4, pp. 453-469, 2002.

[5] H.-S. Ahn and M.E. Lewis, "Flexible server allocation and customer routing policies for two parallel queues when service rates are not additive," *Operations Research*, vol. 61, no. 2, pp. 344-358, 2013.

[6] O.T. Akgun, D.G. Down, and R. Righter, "Energy-aware scheduling on heterogeneous processor," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 599-613, 2014.

[7] S. Andradottir and H. Ayhan, "Throughput maximization for tandem lines with two stations and flexible servers," *Operations Research*, vol. 53, no. 3, pp. 516-531, 2005.

[8] S. Andradottir, H. Ayhan, and D.G. Down, "Server assignment policies for maximizing the steady-state throughput of finite queueing systems," *Management Science*, vol. 47, no. 10, pp. 1421-1439, 2001.

[9] S. Andradottir, H. Ayhan, and D.G. Down, "Dynamic assignment of dedicated and flexible servers in tandem lines," *Probability in the Engineering and Informational Sciences*, vol. 21, no. 4, pp. 497-538, 2007.

[10] S. Andradottir, H. Ayhan, and D.G. Down, "Queueing systems with synergistic servers," *Operations Research*, vol. 59, no. 3, pp. 772-780, 2011.

[11] S. Andradottir, H. Ayhan, and D.G. Down, "Design principles for flexible systems," *Production and Operations Management*, vol. 22, no. 5, pp. 1144-1156, 2013a.

[12] S. Andradottir, H. Ayhan, and D.G. Down, "Optimal assignment of servers to tasks when collaboration is inefficient," *Queueing Systems*, vol. 75, no. 1, pp. 79-110, 2013b.

[13] R. Anupindi and R. Akella, "Diversification under supply uncertainty," *Management Science*, vol. 39, no. 8, pp. 944-963, 1993.

[14] R. Arumugam, M.E. Mayorga, and K.M. Taaffe, "Inventory based allocation policies for flexible servers in serial systems," *Annals of Operations Research*, vol. 172, no. 1, pp. 1-23, 2009.

[15] V. Babich, G. Aydin, P.-Y. Brunet, J. Keppo, and R. Saigal, "Risk, financing and the optimal number of suppliers," in *Supply Chain Disruptions: Theory and Practice of Managing Risks*, H. Gurnani et al. (eds), Springer, 2012.

[16] G.J. Burke, J.E. Carrillo, and A.J. Vakharia, "Sourcing decisions with stochastic supplier reliability and stochastic demand," *Production and Operations Management*, vol. 18, no. 4, pp. 475-484, 2009.

[17] S. Chopra, G. Reinhardt, and U. Mohan, "The importance of decoupling recurrent and disruption risks in a supply chain," *Naval Research Logistics*, vol. 54, no. 5, pp. 544-555, 2007.

[18] F.W. Ciarallo, R. Akella, and T.E. Morton, "A periodic review, production planning model with uncertain capacity and uncertain demand–optimality of extended myopic policies," *Management Science*, vol. 40, no. 3, pp. 320-332, 1994.

[19] M. Dada, N.C. Petruzzi, and L.B. Schwarz, "A newsvendor's procurement problem when suppliers are unreliable," *Manufacturing and Service Operations Management*, vol. 9, no. 1, pp. 9-32, 2007.

[20] L. Dong and B. Tomlin, "Managing disruption risks: the interplay between operations and insurance," *Management Science*, vol. 58, no. 10, pp. 1898-1915, 2012.

[21] A.S. Erdem, M.M. Fadiloglou, and S. Ozekici, "An EOQ model with multiple suppliers and random capacity," *Naval Research Logistics*, vol. 53, no. 1, pp. 101-114, 2006.

[22] T.M. Farrar, *Resource Allocation in Systems of Queues.* Ph.D. Dissertation, Cambridge University, England, 1992.

[23] T.M. Farrar, "Optimal use of an extra server in a two station tandem queueing network," *IEEE Transactions on Automatic Control*, vol. 38, no. 8, pp. 1296-1299, 1993.

[24] A. Federgruen and N. Yang, "Selecting a portfolio of suppliers under demand and supply risks," *Operations Research*, vol. 56, no. 4, pp. 916-936, 2008.

[25] A. Federgruen and N. Yang, "Optimal supply diversification under general supply risks," *Operations Research*, vol. 57, no. 6, pp. 1451-1468, 2009a.

[26] A. Federgruen and N. Yang, "Procurement strategies with unreliable suppliers," *Operations Research*, vol. 59, no. 4, pp. 1033-1039, 2011.

[27] A. Federgruen and N. Yang, "Infinite horizon strategies for replenishment systems with a general pool of suppliers," *Operations Research*, vol. 62, no. 1, pp. 141-159, 2014.

[28] Q. Feng and R. Shi, "Sourcing from multiple suppliers for price-dependent demands," *Production and Operations Management*, vol. 21, no. 3, pp. 547-563, 2012.

[29] E.S. Gel, W.J. Hopp, and M.P. Van Oyen, "Hierarchical cross-training in work-in-process-constrained systems," *IIE Transactions*, vol. 39, no. 2, pp. 125-143, 2007.

[30] B.C. Giri and S. Bardhan, "Coordinating a supply chain under uncertain demand and random yield in presence of supply disruption," *International Journal of Production Research*, vol. 53, no. 16, pp. 5070-5084, 2015.

[31] S. Guo, L. Zhao, and X. Xu, "Impact of supply risks on procurement decisions," *Annals of Operations Research*, vol. 241, no. 1-2, pp. 411-430, 2016.

[32] U. Gurler and M. Parlar, "An inventory problem with two randomly available suppliers," *Operations Research*, vol. 45, no. 6, pp. 904-918, 1997.

[33] J.J. Hasenbein and B. Kim, "Throughput maximization for two station tandem systems: A proof of the Andradottir-Ayhan conjecture," *Queueing Systems*, vol. 67, no. 4, pp. 365-386, 2011.

[34] B. He and Y. Yang, "Mitigating supply risk: An approach with quantity flexibility procurement," *Annals of Operations Research*, vol. 271, no. 2, pp. 599-617, 2018.

[35] W.J. Hopp and M.P. Van Oyen, "Agile workforce evaluation: A framework for cross-training and coordination," *IIE Transactions*, vol. 36, no. 10, pp. 919-940, 2004.

[36] W.J. Hopp, E. Tekin, and M.P. Van Oyen, "Benefits of skill-chaining in production lines with cross-trained workers," *Management Science*, vol. 50, no. 1, pp. 83-98, 2004.

[37] J. Hou and X. Hu, "Comparative studies of three backup contracts under supply disruptions," *Asia-Pacific Journal of Operational Research*, vol. 32, no. 2, 1550006, 2015.

[38] X. Hu, H. Gurnani, and L. Wang, "Managing risk of supply disruptions: incentives for capacity restoration," *Production and Operations Management*, vol. 22, no. 1, pp. 137-150, 2013.

[39] B. Hu and D. Kostamis, "Managing supply disruptions when sourcing from reliable and unreliable suppliers," *Production and Operations Management*, vol. 24, no. 5, pp. 808-820, 2015.

[40] F. Hu, C.-C. Lim, and Z. Lu, "Coordination of supply chains with a flexible ordering policy under yield and demand uncertainty," *International Journal of Production Economics*, vol. 146, no. 2, pp. 686-693, 2013.

[41] H. Huang and H. Xu, "Dual sourcing and backup production: Coexistence versus exclusivity," *Omega*, vol. 57, pp. 22-33, 2015.

[42] K. Jain and E.A. Silver,"The single period procurement problem where dedicated supplier capacity can be reserved," *Naval Research Logistics*, vol. 42, no. 6, pp. 915-934, 1995.

[43] T. Jain and J. Hazra, "Sourcing under incomplete information and negative capacity-cost correlation," *Journal of the Operational Research Society*, vol. 67, no. 3, pp. 437-449, 2016.

[44] E. Kirkizlar, S. Andradottir, and H. Ayhan, "Flexible servers in understaffed tandem lines," *Production and Operations Management*, vol. 21, no. 4, pp. 761-777, 2012.

[45] E. Kirkizlar, S. Andradottir, and H. Ayhan, "Profit maximization in flexible serial queueing networks," *Queueing Systems*, vol. 77, no. 4, pp. 427-464, 2014.

[46] P.R. Kleindorfer and D.J. Wu, "Integrating long- and short-term contracting via business-to-business exchanges for capital-intensive industries," *Management Science*, vol. 49, no. 11, pp. 1597-1615, 2003.

[47] H. Köle and I.S. Bakal, "Value of information through options contract under disruption risk," *Computers and Industrial Engineering*, vol. 103, pp. 85-97, 2017.

[48] G. Koole, "A simple proof of the optimality of a threshold policy in a two-server queueing system," *Systems and Control Letters*, vol. 26, no. 5, pp. 301-303, 1995.

[49] T. Li, S.P. Sethi, and J. Zhang, "Supply diversification with responsive pricing," *Production and Operations Management*, vol. 22, no. 2, pp. 447-458, 2013.

[50] T. Li, S.P. Sethi, and J. Zhang, "Mitigating supply uncertainty: The interplay between diversification and pricing," *Production and Operations Management*, vol. 26, no. 3, pp. 369-388, 2017.

[51] P. Lin and P.R. Kumar, "Optimal control of a queueing system with two heterogeneous servers," *IEEE Transactions on Automatic Control*, vol. 29, no. 8, pp. 696-703, 1984.

[52] E.J. Malecki, "Technology, competitiveness, and flexibility: Constantly evolving concepts," in *The Transition to Flexibility*, D.C. Knudsen (ed), Kluwer, Boston, MA, 1996.

[53] Y. Merzifonluoglou and Y. Feng, "Newsvendor problem with multiple unreliable suppliers," *International Journal of Production Research*, vol. 52, no. 1, pp. 221-242, 2014.

[54] S. Nahmias, *Production and Operations Analysis* (6th ed.). Boston, MA, USA: McGraw-Hill, 2009.

[55] E. Ozkan and J.P. Kharoufeh, "Optimal control of a two-server queueing system with failures," *Probability in the Engineering and Informational Sciences*, vol. 28, no. 4, pp. 489-527, 2014.

[56] D.G. Pandelis, "Optimal use of excess capacity in two interconnected queues," *Mathematical Methods of Operations Research*, vol. 65, no. 1, pp. 179-192, 2007.

[57] D.G. Pandelis, "Optimal stochastic scheduling of two interconnected queues with varying service rates," *Operations Research Letters*, vol. 36, no. 4, pp. 492-495, 2008a.

[58] D.G. Pandelis, "Optimal control of flexible servers in two tandem queues with operating costs," *Probability in the Engineering and Informational Sciences*, vol. 22, no. 1, pp. 107-131, 2008b.

[59] D.G. Pandelis, "Optimal control of noncollaborative servers in two-stage tandem queueing systems," *Naval Research Logistics*, vol. 261, no. 6, pp. 435-446, 2014.

[60] I. Papachristos and D.G. Pandelis, "Optimal dynamic allocation of collaborative servers in two station tandem systems," *IEEE Transactions on Automatic Control*, DOI 10.1109/TAC.2018.2852604, 2018.

[61] M. Parlar and D. Perry, "Inventory models of future supply uncertainty with single and multiple suppliers," *Naval Research Logistics*, vol. 43, no. 2, pp. 191-210, 1996.

[62] M. Parlar and D. Wang, "Diversification under yield randomness in inventory models," *European Journal of Operational Research*, vol. 66, no. 1, pp. 52-64, 1993.

[63] H. Parvin, M.P. Van Oyen, D.G. Pandelis, D.P. Williams, and J. Lee, "Fixed task zone chaining: Worker coordination and zone design for inexpensive cross-training in serial CONWIP lines," *IIE Transactions*, vol. 44, no. 10, pp. 894-914, 2012.

[64] Z. Rosberg, P.R. Varaiya, and J.C. Walrand, "Optimal control of service in tandem queues," *IEEE Transactions on Automatic Control*, vol. 27, no. 3, pp. 600-610, 1982.

[65] Z. Rosberg and A.M. Makowski, "Optimal routing to parallel heterogeneous servers–small arrival rates," *IEEE Transactions on Automatic Control*, vol. 35, no. 7, pp. 789-796, 1990.

[66] S. Saghafian and M.P. Van Oyen, "The value of flexible backup suppliers and disruption risk information: newsvendor analysis with recourse,"*IIE Transactions*, vol. 44, no. 10, pp. 834-867, 2012.

[67] K. Schiefermayr and J. Weichbold, "A complete solution for the optimal stochastic scheduling of a two-stage tandem queue with two flexible servers," *Journal of Applied Probability*, vol. 42, no. 3, pp. 778-796, 2005.

[68] A. Schmitt and L.V. Snyder, "Infinite-horizon models for inventory control under yield uncertainty and disruptions," *Computers and Operations Research*, vol. 39, no. 4, pp. 850-862, 2012.

[69] L.I. Sennott, "The convergence of value iteration in average cost Markov decision chains," *Operations Research Letters*, vol. 19, no. 1, pp. 11-16, 1996.

[70] L.I. Sennott, *Stochastic Dynamic Programming and the Control of Queueing Systems.* New York, N.Y., USA: Wiley-Interscience, 1999.

[71] D.A. Serel, "Capacity reservation with supply uncertainty," *Computers and Operations Research*, vol. 34, no. 4, pp. 1192-1220, 2007.

[72] D.A. Serel, "Optimal resource acquisition policy for a newsvendor under supply risk," *Journal of the Operational Research Society*, vol. 65, no. 11, pp. 1748-1759, 2014.

[73] D.A. Serel, "A single-period stocking and pricing problem involving stochastic emergency supply," *International Journal of Production Economics*, vol. 185, no. C, pp. 180-195, 2017.

[74] R. Serfozo, "An equivalence between continuous and discrete time Markov decision processes," *Operations Research*, vol. 27, no. 3, pp. 616-620, 1979.

[75] F.J. Sting and A. Huchzermeier, "Ensuring responsive capacity: How to contract with backup suppliers," *European Journal of Operational Research*, vol. 207, no. 2, pp. 725-735, 2010.

[76] R.H. Stockbridge, "A martingale approach to the slow server problem," *Journal of Applied Probability*, vol. 28, no. 2, pp. 480-486, 1991.

[77] J.M. Swaminathan and J.G. Shanthikumar, "Supplier diversification: Effect of discrete demand," *Operations Research Letters*, vol. 24, no. 5, pp. 213-221, 1999.

[78] B. Tomlin, "On the value of mitigation and contingency strategies for managing supply chain disruption risks," *Management Science*, vol. 52, no. 5, pp. 639-657, 2006.

[79] B. Tomlin, "Disruption-management strategies for short life-cycle products," *Naval Research Logistics*, vol. 56, no. 4, pp. 318-347, 2009b.

[80] B. Tomlin and Y. Wang, "On the value of mix flexibility and dual sourcing in unreliable newsvendor networks," *Manufacturing and Service Operations Management*, vol. 7, no. 1, pp. 37-57, 2005.

[81] M.P. Van Oyen, E.S. Gel, and W.J. Hopp, "Performance opportunity for workforce agility in collaborative and noncollaborative work systems," *IIE Transactions*, vol. 33, no. 9, pp. 761-777, 2001.

[82] J. Walrand, "A note on "Optimal control of a queueing system with two heterogeneous servers"," *Systems and Control Letters*, vol. 4, no. 3, pp. 131-134, 1984.

[83] X. Wang, S. Andradottir, and H. Ayhan, "Dynamic server assignment with task-dependent server synergy," *IEEE Transactions on Automatic Control*, vol. 60, no. 2, pp. 570-575, 2015.

[84] Y. Wang, W. Gilland, and B. Tomlin, "Mitigating supply risk: Dual sourcing or process improvement?" *Manufacturing & Service Operations Management*, vol. 12 no. 3, pp. 489-510, 2010.

[85] R.R. Weber, "On a conjecture about assigning jobs to processors of different speeds," *IEEE Transactions on Automatic Control*, vol. 38, no. 1, pp. 166-170, 1993.

[86] J. Weichbold and K. Schiefermayr, "The optimal control of a general tandem queue," *Probability in the Engineering and Informational Sciences*, vol. 20, no. 2, pp. 307-327, 2006.

[87] C.-H. Wu, M.E. Lewis, and M. Veatch, "Dynamic allocation of reconfigurable resources in a two-stage tandem queueing system with reliability considerations," *IEEE Transactions on Automatic Control*, vol. 51, no. 2, pp. 309-314, 2006.

[88] C.-H. Wu, D.G. Down, and M.E. Lewis, "Heuristics for allocation of reconfigurable resources in a serial line with reliability considerations," *IIE Transactions*, vol. 40, no. 6, pp. 595-611, 2008.

[89] P. Wu, X. Chao, and J. Chen, "On the replenishment policy considering less expensive but non-committed supply," *Operations Research Letters*, vol. 42, no. 8, pp. 509-513, 2014.

[90] S.H. Xu, "A duality approach to admission and scheduling controls of queues," *Queueing Systems*, vol. 18, no. 3-4, pp. 273-300, 1994.

[91] S. Yang, J. Yang, and L. Abdel-Malek, "Sourcing with random yields and stochastic demand: newsvendor approach," *Computers and Operations Research*, vol. 34, no. 12, pp. 3682-3690, 2007.